

STIC-Biotech/ChemLib

6/1/44

From: Slobodyansky, Elizabeth
Sent: Tuesday, February 26, 2002 5:59 PM
To: STIC-Biotech/ChemLib
Subject: 09/724,126

Please search for case 09/724,126:

SEQ ID NOs: 1 and 2 against commercial ~~and interference~~ databases.

Thank you.

Elizabeth Slobodyansky, PhD

Primary Examiner

Art Unit 1652
CM1 10D11
703-306-3222

mail box 10C01

RECEIVED
FEB 27 2002
STIC

if Contact:
Sheppard

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: _____
Date Completed: 3/2/02
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:

NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST(where applic.)

STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 1, 2002, 12:40:50 ; Search time 95.27 Seconds
(without alignments)
14995.520 Million cell updates/sec

Title: US-09-724-126A-1
Perfect score: 6308
Sequence: 1 gccagaattcgccagcagg.....aatttgatttggtttt 6308

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2.6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2.6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2.6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2.6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2.6/ptodata/2/ina/PCRUS_COMB.seq.*
6: /cgn2.6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4435.4	70.3	6395	2	US-08-982-956-1
2	4435.4	70.3	6395	3	US-09-228-317-1
3	996.2	15.8	1001	2	US-08-982-956-2
4	996.2	15.8	1001	3	US-09-228-317-2
5	57.4	0.9	7218	1	US-08-232-463-14
6	52.2	0.8	7218	1	US-08-232-463-14
7	44.4	0.7	7333	3	US-09-138-024-21
8	44.2	0.7	72928	3	US-09-009-913-1
9	42.2	0.7	325	4	US-08-991-789A-236
10	41	0.6	1517	3	US-08-963-602-6
11	40.4	0.6	152331	3	US-09-128-155-16
12	39.8	0.6	2861	1	US-08-299-953-1
13	39.8	0.6	2861	1	US-08-459-415-1
14	39.8	0.6	2861	5	PCT-US95-11231-1
15	39.8	0.6	3881	1	US-08-299-953-2
16	39.8	0.6	3881	1	US-08-459-415-2
17	39.8	0.6	3881	5	PCT-US95-11231-2
18	39.8	0.6	4376	1	US-08-119-125A-1
19	38.8	0.6	246240	2	US-08-724-394A-20
20	38.8	0.6	246240	2	US-08-724-394A-21
21	38.8	0.6	246240	2	US-08-724-394A-22
22	38.4	0.6	5703	4	US-09-280-590A-36
23	38.4	0.6	18596	4	US-09-318-448-11
24	37.8	0.6	19011	1	US-08-310-356-36
25	37.8	0.6	19557	5	PCT-US92-06300-1
26	37.6	0.6	72928	3	US-09-009-913-1
27	37.4	0.6	1333	4	US-08-543-246B-15

C 28	37.4	0.6	1387	4	US-08-543-246B-1	Sequence 1, Appli
C 29	37	0.6	1113	1	US-08-341-538A-1	Sequence 1, Appli
C 30	37	0.6	1113	2	US-08-725-518-1	Sequence 1, Appli
C 31	37	0.6	1618	3	US-08-889-108-1	Sequence 1, Appli
C 32	37	0.6	1618	3	US-08-889-108-3	Sequence 3, Appli
C 33	37	0.6	1618	4	US-08-120-601B-1	Sequence 1, Appli
C 34	37	0.6	1618	4	US-08-120-601B-3	Sequence 3, Appli
C 35	37	0.6	1618	5	PCT-US94-10358-1	Sequence 1, Appli
C 36	37	0.6	1618	5	PCT-US94-10358-3	Sequence 3, Appli
C 37	37	0.6	4875	1	US-08-460-739-1	Sequence 1, Appli
C 38	36.6	0.6	1480	4	US-09-290-640-65	Sequence 65, Appli
C 39	36.6	0.6	3784	1	US-07-623-033-1	Sequence 1, Appli
C 40	36.4	0.6	1222	4	US-08-543-246B-5	Sequence 5, Appli
C 41	36.2	0.6	9636	1	US-08-323-170B-1	Sequence 1, Appli
C 42	36.2	0.6	19124	2	US-08-487-826B-13	Sequence 13, Appli
C 43	35.4	0.6	252	2	US-08-332-766A-28	Sequence 28, Appli
C 44	35.4	0.6	2135	4	US-08-430-286A-1	Sequence 1, Appli
C 45	35.2	0.6	289	4	US-09-007-005-17	Sequence 17, Appli

ALIGNMENTS

RESULT 1
US-08-982-956-1
; Sequence 1, Application US/08982956
; Patent No. 5861312
; GENERAL INFORMATION:
; APPLICANT: Varshavsky, Alexander
; APPLICANT: Kwon, Yong Tae
; TITLE OF INVENTION: NUCLEIC ACID ENCODING MAMMALIAN UBRL
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kevin M. Farrell, P.C.
; STREET: P.O. Box 999
; CITY: York Harbor
; STATE: ME
; COUNTRY: US
; ZIP: 03911
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/982,956
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Farrell, Kevin M.
; REGISTRATION NUMBER: 35,505
; REFERENCE/DOCKET NUMBER: CIT-2001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (207) 363-0558
; TELEFAX: (207) 363-0528
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6395 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 115..5385
US-08-982-956-1

Query Match 70.3%; Score 4435.4; DB 2; Length 6395;
Best Local Similarity 86.3%; Pred. No. 0;
Matches 4997; Conservative 0; Mismatches 721; Indels 69; Gaps 6;
Oy 590 ttcaaggccgctgtaaaagtctgcctctgtctc-tccgaccggccacaggttttcgct 648

Qy	2809	a	taagttcttgttacttggctacttcagaggtatgaacctgccgagggctttttaacaaaccca	2868
Db	2228	ACAAGTCTTGTTTACTGGTACTTCAGAGATATGAACCTACTGTATGCTTTTAAACAAGACCA	2287	
Qy	2869	tatctcaaaaaacaggattgtttaacaaataatacaactaatagaagaatgcttc	2928	
Db	2288	TAYCCACAAAAGACCAGGATTTGATTAACAGTAGTAATACATTAAATAGAGAATGCTTC	2347	
Qy	2929	aggtcctcatctatattgtgggtgagcgttatgtacctggagtgggaaaagtaccaaag	2988	
Db	2348	AGGTCTCATCTATATTGTGGCAGAACGTTATGTACCTGGAGTGGGAAATGTTACCAGAG	2407	
Qy	2989	agaggtcacaaatgagagaaaatcatcacttgccttgattgaacccatgccacacagt	3048	
Db	2408	AGGAGGTTAATAGAGAGATTACTCACTTACTTTGCATTGCAATGAGCCCATGCCACAGTC	2467	
Qy	3049	ccattgccaaaaatttacctgagaatgaaaataatgaacctggccttagaagaatgtcataa	3108	
Db	2468	CCATGCGCAGAACCTACCTGAGACGAAATAATGAACCTGGCTTTAGAGAAATGTCATAA	2527	
Qy	3109	acaaagtggccacatttaagaaacagggtgatcaggccatggagtttatgaaactaaaag	3168	
Db	2528	ACAAAGTGCCACATTTAAGAAACCAGGTGTCGGGCCATGGAGTTTATGAATTGAAG	2587	
Qy	3169	atgaatcaactgaaagacttcaatatgtactttatcatctccaaacccagcatagca	3228	
Db	2588	ATGAATCACTGAAGACCTTCAATATGTACTTTTACCATTATCTTAAACACACAGCATAGCA	2647	
Qy	3229	aggctgaacatatgcagaagaaaggagaaacaaagaaacaaagatagaagcattgccgc	3288	
Db	2648	AGCTGAACATATGCAAGAAAGAGAGAAACGAAGAAATAAGATGAAGCATTTGCCG	2707	
Qy	3289	caccacacctoctgaattctgcctgcgtcttcagcaagtgattaaacctctcaactgtg	3348	
Db	2708	CGCCACCTCTCCAGAGTTCTGCCCTGCTTTCAGCAAAAGTAGTCAACCTGCTCAGCTGTG	2767	
Qy	3349	atatcatgatgtacattctcaggaccgttatttgagggggcaatatagacagagttctaact	3408	
Db	2768	ATGTTATGATATACATPCTCAGGACCATCTTTGAGCGGGCAGTGGACACGGAGTCTAATC	2827	
Qy	3409	tgtggaccgaaggatgctccaaatggctttcatattcttgcgaattgggttttactagaag	3468	
Db	2828	TGTGGACAGAGGGATGCTGCAGATGGCGTTCATATATTGGCACTGGGCTTCTCTGGAAG	2887	
Qy	3469	agaagcaacagcttcaaaaagcctcgaagaagtagtaacatttgacttttatcataagg	3528	
Db	2888	AGAAGCAGCAGGCTTCAGAAAGCTCCTGGAAGAGGAAGTGGCTTTTGACTTTTACCATAAG	2947	
Qy	3529	cttcaagattgggaagttcagccat-----gaatatataaatgcttttggaaaaaac	3579	
Db	2948	CTTCAAGATTGGGAAGTTCAGGCCATGAATGCTCAGATATATACAATGCTCTTGGAAGAC	3007	
Qy	3580	tcaaaaggaattccccagtttagaaggccagaagacatgaatacgtgtgacttccagatgt	3639	
Db	3008	TCAAAGGAATCCCCAATTAGAAGCCAGAGGACATGATAACATGATCTCCAGATGT	3067	
Qy	3640	tgcacagtgaaagattaaagaaaaatcttgtttaattgttagcaaccacatcaggat	3699	
Db	3068	TTGACACAGTGAAGCGATTAAAGAGAAAAATCTGTTTGTGTTGGCAACCACTTCAGGAC	3127	
Qy	3700	cgaattcttaagaatgatgattaccatgatgaagaaaaagcagcagaacgaagaa	3759	
Db	3128	TGGAGTGCATTAAAGTGTAGGAGATTACTCATGATAAAGAAAGGCGAAGACGGGAAGAA	3187	
Qy	3760	aagctgaagctgctaggtcatcgcgcagaagatacgtgtccagatgtctgccttacaga	3819	
Db	3188	AAGCTGAGGCCGCTAGGCTTCATCGCCAGAAAGATCATGGCCCCAGATGCTGCTCTACAGA	3247	
Qy	3820	aaaacttcattgaaactcaataactcatgtatgacaatatcatcagaaaatgctctgggaag	3879	
Db	3248	AAAACCTCATGAAACCCCAAACTCATGTATGATAATACCTCAGAGTAGTAACAGGAAGG	3307	

Qy	3880	aagattccattatgaggaagagagacccccagcagctcagctgactactactagaaattgctt	3939
Db	3308	AAGACTCCATTATGGAGGAAGAGAGACACCTCAGCAGCTCAGTGTAGAGGCTCTCTAGAATTGCTC	3367
Qy	3940	tgggtccctaaacggggtccatctgtttacttgaagaagagagctgctgacgtcatcctcttggc	3999
Db	3368	TGGGGCCCTTAACGGGGCCCGGTGTTTACCGAAAGAGAGGTGCTGACGTGCATCCTCTGCC	3427
Qy	4000	aagaagaacaggggtgaaaaTagaaaaTaaTgccttggtattatcggcctgtgtccaga	4059
Db	3428	AAGAAGAACAGAGGTAAACTAGAAANATATGCCATGGTATGTCTCAGCATGTGTGCAGA	3487
Qy	4060	aatctatgccttaacccagcacaggggaaaaaccatagaactctcaggagaagccctag	4119
Db	3488	AATCCACCCCTTAACCCAGCAGAGAGGAAGCCTGTGGACCCTTAGGGGAAACACTGG	3547
Qy	4120	accacttttcatggtccagacttggcattatggaacttatacaggaagctgtgggtcatg	4179
Db	3548	ACCCCTCTTTTCATGATCCAGACTTTGGCACATGGAACTTATACAGGAAGCTGTGGTCAATG	3607
Qy	4180	taatcacgcagctgtgctggcagaagtattttgaagctgtcacagctgagcctccagcagc	4239
Db	3608	TAATGCATGCAGTGTCTGGCAGAGATTTTGAAGCTGTGCAGCTGAGCTCGCAGCAGC	3667
Qy	4240	gcattcatgttgacctttttgacttggaaagtggaataattttgccctctttgcaaat	4299
Db	3668	GCATTACAGCTAGACCTGTTTGACCTGGAGAGCGCGAGTACCTATGCCGCTCTGCAAGT	3727
Qy	4300	ctctgtgaataactgtatccccattattctcttggaaacctcaaaagataaacagtgaga	4359
Db	3728	CTCTCTGCAACACTGTCTATCCCCATCATCCCTTTTGCAGCGCGAAGAGATCAACAGTGTAGA	3787
Qy	4360	atgcagatgctcttgcTcaactttgacctggccctggccggtggatcacagactgttctggcca	4419
Db	3788	ATGCGGAGGCTCTTGCTCAACTTTTGACCTTGCCCGGTGGATACAGACTGTCTTGGCCA	3847
Qy	4420	gaatcacgtgtataataaagacatgctaaagagagaaaaacc---caattccctatttct	4476
Db	3848	GAATATCGGGTTATAATATAAAGCATGCTAAAGGAGAAAGGCCAGCAGTTCCTGTCTTGT	3907
Qy	4477	ttaatacaagaatatggagagattctactttggaggttccattccactcctgagttttggcggtg	4536
Db	3908	TTTAATCAAGGAATGGGGGATTCACATTTTGAGTTTTCATTCCATCCCTCAGTTTGGAGTTC	3967
Qy	4537	agctcttcgattaaattcaaatagcatcaaggaataatggttattctcttctggccacaacaa	4596
Db	3968	AGTCTCTCGTGAATAATTCAATATGATATCAAGGAATATGTCATCTCTCGCCACAACAA	4027
Qy	4597	tttatagaattggattgaagtgccacctgatgaagggatcctcagtgccccatgctga	4656
Db	4028	TTTACAGAAATGGCCTGAAAGTGCCCTCCTGATGAACCTAGACCCACAGTGCCCATGATGA	4087
Qy	4657	cctgagacacctgctgttccatctatccaggcaattgaaaaatctatitgggagatgaaggaa	4716
Db	4088	CCTGGAGCAGCTGTGCGTTTCACCATCCAGCAATCGAAAACCTGTGTGGAGATGAAGGAA	4147
Qy	4717	aacctctgtttgagcacctcaaaataggcagcataatggtctggaagcattaatgcagt	4776
Db	4148	AACTCTATTTTGGAGCACTTCAAAATAGACAGCATAGCGGTCTGAAGGCCCTTAATGCAGT	4207
Qy	4777	ttgcagttgcacagagattacctgtcctcaggtccctgatacacgaacaactctggttcgtc	4836
Db	4208	TTGCAAGTTGCACAGAGGGCTACCTGCGCTCAGGTCTGTATACACAACATCTGGCTCGGC	4267
Qy	4837	ttctatcagttgttcttctaatacaataaaatcagaagatacacacatgacctctgtctatag	4896
Db	4268	TCCTGTCTAGTTAATCTTCCCTAAGCTCAATTCAGAAATACACAGGGCTTCTGTCTGTGG	4327
Qy	4897	atctgtttcatgattttgggtgggtgctgtgttagcattcccatcctgtattggatgaacc	4956
Db	4328	ATCTCTTCCATGTTCTGGTGGGGCGAGTCTTAGCGTTCCCATCTCTGTATTTGGATGACA	4387
Qy	4957	ctgtgtgattctgcagccttcttcagtttagtttcttctaataaccaccttattccttcatt	5016


```
Db 4388 CCGTGGATCTGCAGCGCTGCCACTTAGTTCTTCATATAACCACTCTATCTCTCCATC 4447
Qy 5017 tgatccacttgccacactgtcttcagatactacttaacagtagaca-----cag 5064
Db 4448 TGATCACCATGGCGCACATGCTTCAGATACCTCTTACAACAGATACAGATGTGCTCCAG 4507
Qy 5065 gctaccccttgcaggttcaagagacagtagaagagcctcattccgcattcttcttct 5124
Db 4508 GGCGCGCTGTCTGAGGGTGAAGAGATAGTAGGAGGCTCGCTGTGCACTGCTTTCT 4567
Qy 5125 ttgcagaaattcttcaataacacagtggtccattggtggtgatattccctggctggtgatt 5184
Db 4568 TTGTGGAAGTGTGCGACACAGACGGCTCACTAGTGGTGGGTGCTCCCGGCTGGTACC 4627
Qy 5185 tgggtgctactgaagaatggtcatcaccccttatcttcgtgctgctattgtttttcc 5244
Db 4628 TGTGGGCTCTCCCTGAGGAACGGCATCACCCCTTACCTCCGCTGTGCTGCATGCTTTTTC 4687
Qy 5245 actatttacttgggttaactccgctgaggaactgcataccaattctgcagaagagagt 5304
Db 4688 ACTATTACTTGGAGTAGCTCCGCCCTGAAGAAGTGTGTGCAATTCGTGGAAGGAGAA 4747
Qy 5305 acagtgcactctgactatctatctttaccatacaaaattgttccctctctccaggaa 5364
Db 4748 TCAGTGCACCTCTAGTATCTATCTTACCACAAATTTGTTCTCTGCTTTTCCAGGAAT 4807
Qy 5365 attgggatactgtaagcccttgcctccagaggtggtggtgcagatccctgccttaaaact 5424
Db 4808 ATTGGGATACCATAAAGCCCTTACTACAGAGGTGGTGGGAGATCCCGCTTACTCAAGT 4867
Qy 5425 gtttgaagcaaaaaacacagtggtcaggtaccctagaaagaaataagtttgatagagc 5484
Db 4868 CTTTGAAGCAGAAAAGTGTGTGGTGGTACCCCTAGAAAAGAAATAGTTTGATAGAGC 4927
Qy 5485 ttctgtgactatagctgctcctcctgaatcaagctctcattcaggtgccacggctg 5544
Db 4928 TTCTCGAGACTACAGCTGTCTTCTAAATCAGGCTTCTCACTTAGGTGTCCACGGTCTG 4987
Qy 5545 cagatgatagcgaaagcactctgctcctctgcttcttctgtggtggtatatactattctc 5604
Db 4988 CAGATGATGAGCAAGACATCTCTCTCTGTCTTTTCTGTGGGCCATCTGTGTTCTC 5047
Qy 5605 agaacttctgctccaggaaaattgtgaacggggaagaggtgagcttgcatttttcaag 5664
Db 5048 AGAACATCTGTTGCCAAGAAATAGTGAATGGGAAGAGGTTGAGCGTGGCTTTTCATG 5107
Qy 5665 cactcactgtgagcgcgagctgctgatttctcctaaatacagaaatccgagtggtcc 5724
Db 5108 CGTTTCATTTGTGTGGAGTCTGCATTTTCTTAAATCCGAGATCGAGGTTGGTCC 5167
Qy 5725 tgggtgaaggttaaacgaggtgctgctctatccagctcctccttcttgatgaatatgag 5784
Db 5168 TGGTGAAGGAAAAGCCAGAGGCTGTGCTTACCAGCCCTTACTTGGATGAATATGAG 5227
Qy 5785 aaacagacctggcctgaagagaggggcaacccctcatttctatctgtagcggtagcgga 5844
Db 5228 AAACAGACCCAGGGCTTAAAGAGAGAAACCCACTTCAATTATCTCGGAGCGGTATCGGA 5287
Qy 5845 agctccatttggctgagcaaacactgcattatagaagagattgctagagcgaagaga 5904
Db 5288 AGCTGCATTTGGTCTGGCAACACATGTCATATAGAAGAGATTGCTCGGAGCCAGAGA 5347
Qy 5905 ctaatcagatgttatttgattcaactggcagttactgtgagctccaaactctgcctcaag 5964
Db 5348 CTAATCAGATGCTATTGGATTAACTGGCAGTTACTCTGAGGCTTCAGTTCTGCCCTCAAG 5407
Qy 5965 acaatcacaaatgacacagtagtaagagctgattcaaaattatggaaaaactttctgag 6024
Db 5408 ACAATCATGAGTCACATCAATAATAAGACTGATCTAAAATTTCTAGAGAACTTCTGAGG 5467
Qy 6025 gctgggaaagattgaggggtcttttgcctcatgtccaggttccacttaacataaaat 6084
```

```
Db 5468 ACGGGGAAGTATTGGAGGTCTTTTGATCCATGTCCAGATTCCACACATTATATAAAT 5527
Qy 6085 atttcttaatgagattgtcttcaattagcaaacatatgtcttcacaggaaaaa-aggac 6143
Db 5528 ATTCTTATGGAATATTGCTTTCAATTATCAAAACATAAGCTTCAAGGGAAAAACAAGAC 5587
Qy 6144 atagatcaatctgt-----ttt 6160
Db 5588 ATAGATTAAATGTTTATGTTTAGAACACATAAGAAATGCTTGTTCATCCAAAGTGTCTAT 5647
Qy 6161 atgtctctagtagtattccaggaaatttattcccttccataattgtctcatttcttattt 6220
Db 5648 TTCTCTAATATTTCAGAAAACTCTTTCCCTTCATAACTGCTTAGTTCAATTTTATAT 5707
Qy 6221 catccacttgtagatgaagtccacgtcaaacagttgtagacattttatgtgttgggttaac 6280
Db 5708 CACCACCTGGTAAATAGGTCACATTAAGCATTTTGTGGACATTTCTCCATCTGGGTAAAC 5767
Qy 6281 tctctgcaatttctatttgggtttt 6307
Db 5768 ATCTCTGCACCTTTGTATTGTTGGTGT 5794

RESULT 2
US-09-228-317-1
; Sequence 1, Application US/09228317
; Patent No. 6159732
; GENERAL INFORMATION:
; APPLICANT: Varshavsky, Alexander
; APPLICANT: Kwon, Yong Tae
; TITLE OF INVENTION: NUCLEIC ACID ENCODING MAMMALIAN UBRI
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kevin M. Farrell, P. C.
; STREET: P.O. Box 999
; CITY: York Harbor
; STATE: ME
; COUNTRY: US
; ZIP: 03911
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/228,317
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Farrell, Kevin M.
; REGISTRATION NUMBER: 35,505
; REFERENCE/DOCKET NUMBER: CIT-2001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (207) 363-0558
; TELEFAX: (207) 363-0528
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6395 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 115..5385
US-09-228-317-1
```

Query Match 70.3%; Score 4435.4; DB 3; Length 6395;
Best Local Similarity 86.3%; Pred. No. 0;
Matches 4997; Conservative 0; Mismatches 721; Indels 69; Gaps 6;

Qy 590 ttcaggggcccgtcgtaaaagtgtcgtccctgtctc-tccgaccgcccacaggtttccgct 648

Db	4388	CCGTGATCTGCAGCGTGGCCACTTAGTCTTCATATAACCACTCTATCTCTTCATC	4447
Qy	5017	tgatcaccattggcacacatggtcttaagatctactacttaacagtagaca-----cag	5064
Db	4448	TGATCACATGGCGCACATGCTTCAGATACCTCTTACACAGATACAGATCTGTCTCCAG	4507
Qy	5065	gcctaccccttgtctcaggttccaagaagacagtgaaagagctcatctccgcatctctttct	5124
Db	4508	GGCGCGGCTTGCTGAGGGTGAAGAGTATGTAGAGGAGCTCGCTGTGCATCTGCTTTCT	4567
Qy	5125	ttcagaagaattctcaatatacaagtggtccattgggtgtgatatctcgtgctggtatt	5184
Db	4568	TTGTGAAGTGTCCGACACACAGACGGCCCTACTGGGTGGGGTGTCTCCGGGTGTGTAAC	4627
Qy	5185	tgtgggtctcactgaagaatggcatcaccccttactctgctgtgctgcatgtttttcc	5244
Db	4628	TGTGGCTCTCCCTGAGGAACGGCATCACCCCTTACCTCCGCTGTCTGTGCACCTGCTTTTCC	4687
Qy	5245	actatttacttgggttaactccgctggaggaactgcataccaattctgcagaaggagagt	5304
Db	4688	ACTATTTACTTGGAGTAGCTCGCCTGAAGAACTGTTTGCCAATTCTGCTGAAGAGAAT	4747
Qy	5305	acagtgcaacttgttagctatctatttacctacaaatttgttctcgtctcttcaggaaat	5364
Db	4748	TCAGTGCACCTGTAGCTATCTATCTTTACCCACAAATTTTGTCTCTGCTTTTCCAGGAAT	4807
Qy	5365	attgggatactgaaggcccttgctccagagtggtgtgcagatccctgccttactaaact	5424
Db	4808	ATTGGGATACCATATAGGCCCTTACTACAGAGGTGTGTGGAGATCTCTGCCTTACTCAAGT	4867
Qy	5425	gtttgaagcaaaaaacaccgtggtcaggtacccttagaaaaaagaatagtttgatagagc	5484
Db	4868	CTTTGAAGCAGAAAAGTGTGTGGTCAAGTACCCTAGACCTAGAAAAAGAAATAGTTGATAGAGC	4927
Qy	5485	ttcctgatgactatagctgctcctcctgaatcaagcttctcatctcaggtgccacggtctg	5544
Db	4928	TTCTCTGAGACTACAGCTGTCTTCTAAATCAGGCTTCTCACTTTAGGTGTCCAGGGTCTG	4987
Qy	5545	cagatgatgacgaagaacatcctctcctcgtccttctgtggggtctatactatgtctc	5604
Db	4988	CAGATGATCAGGCGAAGACATCCTGTCTCTGTCTTTCTGTGGGGCCATCCTGTGTCTC	5047
Qy	5605	agaacatttgtccagaaaaatttgaaagggaagggttggagcttgcaattttcaag	5664
Db	5048	AGAACATCTGTTGCCAAGAAATAGTGAATGGGAAGAGTTGGAGCGTGGCTTTTTCATG	5107
Qy	5665	caattcaattgtgagcccgagctgcattttcctataaaatcagagaatgcgagtgctcc	5724
Db	5108	CGCTTCATTGTGGTGTCTGGAGTCTGCATTTTTCCTTAAATCCGAGAATGACGGGTGTCTC	5167
Qy	5725	tgttgaaggtaaagccagaggtgtgcctaccagctcccttacttgatgaatatgag	5784
Db	5168	TGTTGAAGGAAGAACCCAGAGGCTGTGCTTACCCAGCCCCCTTACTTGAATGAATATGGAG	5227
Qy	5785	aaacagacctggcctgaagaggggcaaccccttcatattatcgtgagcggatcogga	5844
Db	5228	AAACAGACCCAGGGCTAAAGAGAGGAAACCCACATTCATTTATCTCGGGAGGGTATCGGA	5287
Qy	5845	agctccatttgtctggcaacaaactgcataatagaagaaattgctagagaccaagaga	5904
Db	5288	AGCTGCATTTTGGTCTGGCAACAGCACTGCATTTATAGAAAGATTTGCTCGGAGCCAGGAGA	5347
Qy	5905	ctaatacagatgtatttgattcaactgagcgagttactgtgagctcccaactctgcctcaag	5964
Db	5348	CTAATCAGATGCTATTTGATTTAACTGGCAGTTACTCTGAGCTTCTAGTTCTTGCTCAAG	5407
Qy	5965	acaatacacaatlgacgacagtagtaaggctgattcacaattatggaaaaacttctcaggt	6024
Db	5408	ACAATCATGAGTGACATCAATAATAAAGACATGATCTAAATTTCTAGAGAACTTTTCTGAGG	5467
Qy	6025	gctggaaagtattggaggggtttttgtccatgtcccaagttcacttacctacataaaaaat	6084

```

Db 5468 ACGGGGAAGTATTGGAGGCTCTTTTGATCCATGTCCACACATTTAATAAAAT 5527
Qy 6085 atttcttaagtgattgtcttccaattagcaacacatagtcttcacaggaaaaa-aggac 6143
Db 5528 ATTCCTTAATGGAATATTGCTTTCAATTATCAAAACATAAGCTTCAAGGGGAAAAACAAGAC 5587
Qy 6144 atagatcaatctgt-----ttt 6160
Dy 5588 ATGAGTAAATGTTTATGTTCTAGAACACTAAAGAAATGCTGTGTTCAATCCAAAGTGCTAT 5647
Qy 6161 atgtgctagattccagggaatttatccccccttcataaattgtctcatttcattttattt 6220
Dy 5648 TTCTGCTATATATTCAGAAAACTCTTTCCCTTCATAACTGTCCTAGTTCAATTTCATAT 5707
Qy 6221 catccactggtagtagaagtcacgcacgtcaaacagttgtgacacattttatggtggttaac 6280
Dy 5708 CACCACACCTGGTTAATGAGGTACATTAAGCATTTGTGGACATTTCTCCATCTGGCTAAC 5767
Qy 6281 tcttctgaattttatttgtgttt 6307
Dy 5768 ATCTGTCACCTTGTATTGGTGTTT 5794

RESULT 3
US-08-982-956-2
; Sequence 2, Application US/08982956
; Patent No. 5861312
; GENERAL INFORMATION:
; APPLICANT: Varshavsky, Alexander
; APPLICANT: Kwon, Yong Tae
; TITLE OF INVENTION: NUCLEIC ACID ENCODING MAMMALIAN UBRI
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kevin M. Farrell, P.C.
; STREET: P.O. Box 999
; CITY: York Harbor
; STATE: ME
; COUNTRY: US
; ZIP: 03911
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/982,956
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Farrell, Kevin M.
; REGISTRATION NUMBER: 35,505
; REFERENCE/DOCKET NUMBER: CIT-2001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (207) 363-0558
; TELEFAX: (207) 363-0528
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1001 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..999
US-08-982-956-2

```

Query Match 15.8%; Score 996.2; DB 2; Length 1001;
Best Local Similarity 99.7%; Pred. No. 7.6e-279;
Matches 998: Conservative 0; Mismatches 3; Indels 0;

Qy 2799 atggatcccaataagttcttgttactggtactgtcagaggtagaacttgccgaggctttt 2858


```

; TELEFAX: 650-327-3231
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 72928 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
US-09-009-913-1

Query Match          0.7%; Score 44.2; DB 3; Length 72928;
Best Local Similarity 86.0%; Pred. No. 0.18;
Matches 49; Conservative 0; Mismatches 8; Indels 0; Gaps

QY 303 cccctgaacaccattcactttctgtctatgaatttgacctactctagctgagt 359
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |
Db 19091 CCCGCGCAACCACCATCTACTTCTCTCTATGAATTGGACCACTCTAAGTAGCT 19035

RESULT 9
US-08-991-789A-236/C
; Sequence 236, Application US/08991789A
; Patent No. 6225054
; GENERAL INFORMATION:
; APPLICANT: Fridakis, Tony N.
; Smith, John M.
; Reed, Steven G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF BREAST CANCER
; NUMBER OF SEQUENCES: 292
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed IP Law Group
; STREET: 701 Fifth Avenue, Suite 6300
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatenLin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/991,789A
; FILING DATE: 11-Dec-1997
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Potter, Jane E. R.
; REGISTRATION NUMBER: 33,332
; REFERENCE/DOCKET NUMBER: 210121.419C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 236:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 325 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; SEQUENCE DESCRIPTION: SEQ ID NO: 236:
US-08-991-789A-236

Query Match          0.7%; Score 42.2; DB 4; Length 325;
Best Local Similarity 79.4%; Pred. No. 0.021;
Matches 50; Conservative 0; Mismatches 13; Indels 0; Gaps
QY 298 acacacctggaaaccaccattctacttttgtctgtatgaatttgcactacttagctgg 351

```

Db	283	AAAGCCCTGGCAACACCATCTATTGTGTCTATGAATTCACCTACTAGGTAC	224
Qy	358	atc 360	
Db	223	CTC 221	
RESULT 10			
US-08-963-602-6/c			
; Sequence 6, Application US/08963602			
; Patent No. 6090554			
; GENERAL INFORMATION:			
; APPLICANT: Woychik, Richard			
; APPLICANT: Garfinkel, David			
; TITLE OF INVENTION: EFFICIENT CONSTRUCTION OF GENE			
; TITLE OF INVENTION: TARGETING VECTORS			
; NUMBER OF SEQUENCES: 8			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &			
; STREET: 233 South Wacker Drive/6300 Sears Tower			
; CITY: Chicago			
; STATE: Illinois			
; COUNTRY: United States of America			
; ZIP: 60606-6402			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: Floppy disk			
; COMPUTER: IBM PC compatible			
; OPERATING SYSTEM: PC-DOS/MS-DOS			
; SOFTWARE: PatentIn Release #1.0, Version #1.30			
; CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: US/08/963,602			
; FILING DATE:			
; CLASSIFICATION: 800			
; ATTORNEY/AGENT INFORMATION:			
; NAME: Clough, David W.			
; REGISTRATION NUMBER: 36,107			
; REFERENCE/DOCKET NUMBER: 01017/33985			
; TELECOMMUNICATION INFORMATION:			
; TELEPHONE: (312) 474-6300			
; TELEFAX: (312) 474-0448			
; INFORMATION FOR SEQ ID NO: 6:			
; SEQUENCE CHARACTERISTICS:			
; LENGTH: 1517 base pairs			
; TYPE: nucleic acid			
; STRANDEDNESS: single			
; TOPOLOGY: linear			
; MOLECULE TYPE: other nucleic acid			
; DESCRIPTION: /desc = "primer"			
US-08-963-602-6			
Query Match 0.6%; Score 41; DB 3; Length 1517;			
Best Local Similarity 56.2%; Pred. No. 0.13;			
Matches 77; Conservative 0; Mismatches 60; Indels 0; Gaps			
Qy	109	gagaaacttacctcaagcgttatcatctgtttctccacttatccacaagctgctatgac	168
Db	1036	GTGTGCTGCTCTCAGAAAGCCGTCTCTTCTGGGGCGCTGGTACCGCGCTGCTGCTGCTGC	977
Qy	169	cactgcctgtgcccgcacctggaaaccaccgcagagcccccactactgcctccaccactgag	228
Db	976	AGTCCCGCTGCCCGCGCTCTGCCACCACCAATTGTACTGTACGCGCGCTCCACCACCTGC	917
Qy	229	ttctccaccctgatca 245	
Db	916	TGCTCTCTCTCTCCACCA 900	
RESULT 11			
US-09-128-155-16			
; Sequence 16, Application US/09128155			
; Patent No. 6117654			


```

US-08-299-953-1

Query Match          0.6%; Score 39.8; DB 1; Length 2861;
Best Local Similarity 46.5%; Pred. No. 0.42;
Matches 128; Conservative 0; Mismatches 147; Indels 0; Gaps

Qy 3607 agaagcagatgaacgtggatacttcagatgtttgcacacagtgaaagcgattaaagagaaa 3666
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1290 AAAAAAGACGGAGGGAGTTAAATATTTTAGATTTTTTACACGTATTAAAAAATATATATCAA 1231

Qy 3667 aatctgtttaatttagcaacacacatcaggatcggaatctattaaagaatgatgagatta 3726
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1230 AAAAAATATCTTTTATATATTAATAAAAAATGGAGAAGATAAATTTATATAATTAATAAAAAAAG 1171

Qy 3727 ctcatgataaagaaaaagcagaaacgaaaaaagaaagctgaagctgctaggctacatcgcc 3786
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1170 ATAAATAAATAAAATCTAGAGTTATAATAAAAAACTAATAATTAATCTTTTTTAACATTGTA 1111

Qy 3787 agaagatcatggctcagatgtctgccttacagaaaaaacttcattgaaactcataaactca 3846
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1110 AAATCATTTATATTATGATATAATTTTTTTCAAACAACCATTAATAAAAAATGATAGGG 1051

Qy 3847 tgtatgaacaatacatcagaaaatgcctgggaagaaa 3881
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1050 AGTATTATCATATGTGCAGAAATATTATATAAAGAA 1016

RESULT 13
US-08-459-415-1/c
; Sequence 1, Application US/08459415
; Patent No. 5744334
; GENERAL INFORMATION:
; APPLICANT: Dobres, Michael S. and Mandaci, Sevnrur
; TITLE OF INVENTION: A Plant Promoter Useful for Directing the
; TITLE OF INVENTION: Expression of Foreign Proteins to the Plant Epide
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5744334ris
; STREET: One Liberty Place 46th. Floor
; CITY: Philadelphia
; STATE: PA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/459,415
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/299,953
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Beardell, Lori Y.
; REGISTRATION NUMBER: 34,293
; REFERENCE/DOCKET NUMBER: NOVA-0003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-564-8960
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2861 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-459-415-1

```

Query Match 0.6%; Score 39.8; DB 1; Length 2861;
Best Local Similarity 46.5%; Pred. No. 0.42;
Matches 128; Conservative 0; Mismatches 147; Indels 0; Gaps 0;
Qy 3607 agaagacatgaacgtggaatcagatggtttgacacagtggaagcgattaaagagaaa 3666
Db 1290 AAAAAGAACGGAGGAGTAAATTTAGATTTTACACGATTAAAAAATTATATCAA 1231
Qy 3667 aatctgttttaattgtagcaaccacatcagatcggaatctattaagaatgatgagatta 3726
Db 1230 AAAAATATCTTTATATATTAATAAATGGAGAGATAAATTTATATAATTAATAAATAAAG 1171
Qy 3727 ctcatgataaagaaagcagaagaaagaaagcgaagcgtgagcgtgagcgtacatcgcc 3786
Db 1170 ATAATAATAAATCTAGAGTTAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1111
Qy 3787 agaagacatggtcagatgctcgtccttacagaaaaactcattgaaactcataaactca 3846
Db 1110 AAATGATTATATATGATATAATTTTTTCAAAACCAACCATATAAATAAATGATAGG 1051
Qy 3847 tgatgacaatcacatcagaaatgcctggaagaaa 3881
Db 1050 AGTATTATCATATGTCAGAAATTTATTATAAAGAA 1016

RESULT 14
PCT-US95-11231-1/c
; Sequence 1, Application PC/TUS9511231
; GENERAL INFORMATION:
; APPLICANT: Dobres, Michael S. and Mandaci, Sevnuur
; TITLE OF INVENTION: A Plant Promoter Useful for Directing the Expression
; OF FOREIGN PROTEINS TO THE PLANT EPIDERMIS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & Norris
; STREET: One Liberty Place 46th. Floor
; CITY: Philadelphia
; STATE: PA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/11231
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 08/299,953
; FILING DATE: September 2, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Beardell, Lori Y.
; REGISTRATION NUMBER: 34,293
; REFERENCE/DOCKET NUMBER: NOVA-0016
; TELEPHONE: 215-564-8960
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2861 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
PCT-US95-11231-1

Query Match 0.6%; Score 39.8; DB 5; Length 2861;
Best Local Similarity 46.5%; Pred. No. 0.42;
Matches 128; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

Matches 128; Conservative 0; Mismatches 147; Indels 0; Gaps 0;
Qy 3607 agaagacatgaacgtggaatcagatggtttgacacagtggaagcgattaaagagaaa 3666
Db 1290 AAAAAGAACGGAGGAGTAAATTTAGATTTTACACGATTAAAAAATTATATCAA 1231
Qy 3667 aatctgttttaattgtagcaaccacatcagatcggaatctattaagaatgatgagatta 3726
Db 1230 AAAAATATCTTTATATATTAATAAATGGAGAGATAAATTTATATAATTAATAAATAAAG 1171
Qy 3727 ctcatgataaagaaagcagaagaaagaaagcgaagcgtgagcgtgagcgtacatcgcc 3786
Db 1170 ATAATAATAAATCTAGAGTTAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1111
Qy 3787 agaagacatggtcagatgctcgtccttacagaaaaactcattgaaactcataaactca 3846
Db 1110 AAATGATTATATATGATATAATTTTTTCAAAACCAACCATATAAATAAATGATAGG 1051
Qy 3847 tgatgacaatcacatcagaaatgcctggaagaaa 3881
Db 1050 AGTATTATCATATGTCAGAAATTTATTATAAAGAA 1016

RESULT 15
US-08-299-953-2/c
; Sequence 2, Application US/08299953
; Patent No. 5646333
; GENERAL INFORMATION:
; APPLICANT: Dobres, Michael S. and Mandaci, Sevnuur
; TITLE OF INVENTION: A Plant Promoter Useful for Directing the
; TITLE OF INVENTION: Expression of Foreign Proteins to the Plant Epidermis
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5646333ris
; STREET: One Liberty Place 46th. Floor
; CITY: Philadelphia
; STATE: PA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/299,953
; FILING DATE: Herewith
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Beardell, Lori Y.
; REGISTRATION NUMBER: 34,293
; REFERENCE/DOCKET NUMBER: NOVA-0003
; TELEPHONE: 215-564-8960
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3881 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-299-953-2

Query Match 0.6%; Score 39.8; DB 1; Length 3881;
Best Local Similarity 46.5%; Pred. No. 0.51;
Matches 128; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

Qy 3607 agaagacatgaacgtggaatcagatggtttgacacagtggaagcgattaaagagaaa 3666
Db 1290 AAAAAGAACGGAGGAGTAAATTTAGATTTTACACGATTAAAAAATTATATCAA 1231

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 1, 2002, 12:32:35 ; Search time 2649.68 Seconds
(without alignments)
25582.097 Million cell updates/sec

Title: US-09-724-126A-1
Perfect score: 6308
Sequence: 1 gccagaattgcgcagagg.....aatttattgttggtttt 6308

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues
Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_estl:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pin:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	674.6	10.7	682	10	AI929033	AI929033 au64c10.y
2	656.6	10.4	756	11	BG334574	BG334574 602533425
C 3	637.4	10.1	641	10	AI361043	AI361043 qy03f11.x
4	627.2	9.9	797	11	BG862813	BG862813 602799074
5	573.2	9.1	782	11	BI086469	BI086469 602849734
C 6	557	8.8	565	11	BF063405	BF063405 7b89d06.x
7	512.2	8.1	606	10	BE589438	BE589438 195602 BA
C 8	479.8	7.6	624	11	BG219270	BG219270 RST39023
C 9	477.8	7.6	481	10	AI192195	AI192195 qc92608.x
10	473.4	7.5	583	10	AW971391	AW971391 EST383480
11	471.6	7.5	478	10	AA401319	AA401319 zu63d04.r
C 12	454	7.2	522	10	AW291190	AW291190 UI-H-BI2-

13	449	7.1	522	11	BG382648	BG382648 298622 MA
14	448	7.1	520	11	BG382624	BG382624 298590 MA
C 15	410.8	6.5	486	10	BE650873	BE650873 UI-M-BH3-
16	354.8	5.6	925	10	BE573578	BE573578 601333605
C 17	352	5.6	374	10	AA400279	AA400279 zu63d04.s
18	351.4	5.6	435	10	AW489271	AW489271 UI-M-BH3-
19	330	5.2	394	11	BF774117	BF774117 283680 MA
C 20	318	5.0	481	10	AI187306	AI187306 qf28h08.x
21	307.2	4.9	397	10	AW311960	AW311960 6005 MARC
22	301.8	4.8	504	10	AI980640	AI980640 pat_pk002
23	266.2	4.2	327	10	BE654236	BE654236 UI-M-AKI-
24	265.2	4.2	523	11	BG797647	BG797647 lcl16b02.x
C 25	263.4	4.2	300	11	Z17892	Z17892 HSDH1041 S
26	252.6	4.0	1079	11	BG292980	BG292980 602389655
C 27	249	3.9	562	10	AI504731	AI504731 v113c11.x
28	243.6	3.9	372	11	BF542537	BF542537 UI-R-C3-s
29	242.8	3.8	505	11	BG364916	BG364916 dc93c08.y
30	241	3.8	325	10	AI615529	AI615529 v113c11.y
31	237	3.8	291	11	AV225341	AV225341 AV225341
C 32	235.8	3.7	396	11	BE930958	BE930958 RC3-GN007
33	232.8	3.7	256	10	AI921294	AI921294 wo23a03.x
C 34	231.4	3.7	278	10	BE077143	BE077143 RC5-BT060
35	230.6	3.7	323	11	H33916	H33916 EST110358 R
C 36	223.6	3.5	745	10	AW976158	AW976158 EST388267
37	222	3.5	555	11	BE873236	BE873236 601451771
38	217	3.4	685	10	AI693180	AI693180 wd68e01.x
39	209	3.3	313	10	AV168252	AV168252 AV168252
40	208.6	3.3	632	11	BG625558	BG625558 pgn1c.pk0
41	206.8	3.3	221	10	AA507138	AA507138 nh42d07.s
42	204	3.2	1148	11	BF164318	BF164318 601773090
C 43	201	3.2	620	10	AW702134	AW702134 uq98d05.x
44	200.4	3.2	435	11	BE930879	BE930879 RC3-GN007
C 45	200	3.2	711	10	AI646734	AI646734 ub65b06.x

ALIGNMENTS

RESULT 1	AI929033	682 bp	EST	23-AUG-1999
LOCUS	au64c10.y1	Schneider fetal brain 00004 Homo sapiens cDNA clone		
DEFINITION	IMAGE:2519538 5' similar to TR:070481 070481 UBIQUITIN-PROTEIN			
	LIGASE E3 COMPONENT N-RECOGNIN ;, mRNA sequence.			
ACCESSION	AI929033			
VERSION	AI929033.1	GI:5664997		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	Hillier, L., Allen, M., Bowles, L., Dubouque, T., Geisels, G., Jost, S., Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.			
TITLE	WashU-NCI human EST Project			
JOURNAL	Unpublished (1997)			
COMMENT	Other_ESTS: au64c10.x1			
	Contact: Wilson RK			
	Washington University School of Medicine			
	444 Forest Park Parkway, Box 8501, St. Louis, MO 63108			
	Tel: 314 286 1800			
	Fax: 314 286 1810			
	Email: est@watson.wustl.edu			
	This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.			
	Possible reversed clone: similarity on wrong strand			
	Seq primer: -40RP from Gibco			
	High quality sequence stop: 450.			
FEATURES	Location/Qualifiers			
source	1..682			
	/organism="Homo sapiens"			

```

BASE COUNT      274 a   138 c   152 g   192 t
ORIGIN
          CA). NOTE: THIS IS A MIRA-MC LIBRARY.

Query Match           10.4%; Score 656.6; DB 11; Length 756;
Best Local Similarity 95.9%; Pred. No. 6.5e-143;
Matches 730; Conservative 0; Mismatches 19; Indels 12; Gaps

QY    2836 ggatcaacttgcaggcgtttttaacagaccatatctacaaagaccaggatttgattta 2895
      |||||
Db     1 GGATGAACCTGCCGAGGCTTTTAAACAAGACCATACTCAAAAGACCAGGATTGTGATTA 60

QY    2896 acaataataacactaatagaagaagtcttcaggctcctcatctatatgtggggtgagc 2955
      +|||||
Db     61 AACAAATATATACACTAATAGAGAANAATGCTTCAGGTCCCTCATCTATATATTGGGGTGAGC 120

QY    2956 gttagtaccttgagtgggaaaattgtgaccaagaagaaggtcccaaagagaaaaatcattc 3015
      |||||
Db     121 GTTATGCTACCTGGAGTGGGAAAATGTGCACCAAGAACAGGCTCAATAGAGAAAAATCATTC 180

QY    3016 acttgctttgcattgaacctatgccacagtygccattgccaaaaatttaccctgagaaatcg 3075

Db     181 ACPTTGCTTTGCTATTGAACCCATTGCCACAGTGCCTTGC AAAAATTTTACCTGAGAAATG 240

```

Qy	3076	aaataatgaactgcttagaagaatgctataaacaagaatggtccacatttaagaacaacag	3135
Db <td>241</td> <td>AAAAATATGAATGGCTTAGACAAATGTCATAAACAAAGTGGCCACATTTAAGAAACACG</td> <td>300</td>	241	AAAAATATGAATGGCTTAGACAAATGTCATAAACAAAGTGGCCACATTTAAGAAACACG	300
Qy <td>3136</td> <td>gtgtatcagccatgaggttatgaactaaaaatgaatcaactgaagacttcaatatgt</td> <td>3195</td>	3136	gtgtatcagccatgaggttatgaactaaaaatgaatcaactgaagacttcaatatgt	3195
Db <td>301</td> <td>GTGTATCAGGCCATGGAGTTTATGAACATAAAGATGAATCAGTGAAGACATTCATATGT</td> <td>360</td>	301	GTGTATCAGGCCATGGAGTTTATGAACATAAAGATGAATCAGTGAAGACATTCATATGT	360
Qy <td>3196</td> <td>actttatcatctactcaaaacccagcagcaaggtgtaacatatagcagaagaaaaagga</td> <td>3255</td>	3196	actttatcatctactcaaaacccagcagcaaggtgtaacatatagcagaagaaaaagga	3255
Db <td>361</td> <td>ACTTTTATCATTAACCAAAACCCAGCATAGCAAGGCTGAACATATCCAGAGAAAGGA</td> <td>420</td>	361	ACTTTTATCATTAACCAAAACCCAGCATAGCAAGGCTGAACATATCCAGAGAAAGGA	420
Qy <td>3256</td> <td>gaaacaagaataaagaatgaag-----catggccgacacacacacctcctgaattct</td> <td>3309</td>	3256	gaaacaagaataaagaatgaag-----catggccgacacacacacctcctgaattct	3309
Db <td>421</td> <td>GAAGAACAGAAACAAAGATGAAGGTAAACAAATGTCGCGCACACACACCTCTCTGAATCT</td> <td>480</td>	421	GAAGAACAGAAACAAAGATGAAGGTAAACAAATGTCGCGCACACACACCTCTCTGAATCT	480
Qy <td>3310</td> <td>gcctgtcttcacaaagtgattaaacctctcaactgtgat-catgatgtacattctc</td> <td>3368</td>	3310	gcctgtcttcacaaagtgattaaacctctcaactgtgat-catgatgtacattctc	3368
Db <td>481</td> <td>GCCCTGCTTTCAGCAAAAGTGATTAACCTTCTCAACTGTGATATCCATGATGTACATTCTC</td> <td>540</td>	481	GCCCTGCTTTCAGCAAAAGTGATTAACCTTCTCAACTGTGATATCCATGATGTACATTCTC	540
Qy <td>3369</td> <td>aggacctatttgagcgggcaatagacacagattctaaacttggaccgaaggatgctc</td> <td>3428</td>	3369	aggacctatttgagcgggcaatagacacagattctaaacttggaccgaaggatgctc	3428
Db <td>541</td> <td>AGGACCGTATTGACCGGGCAATAGACACAGATTTCAACTTGTGGACCAAGGATGCTC</td> <td>600</td>	541	AGGACCGTATTGACCGGGCAATAGACACAGATTTCAACTTGTGGACCAAGGATGCTC	600
Qy <td>3429</td> <td>caaatggcttttcatacttctggcattgggtttactagaagagagcaacagcttcaaaa</td> <td>3488</td>	3429	caaatggcttttcatacttctggcattgggtttactagaagagagcaacagcttcaaaa	3488
Db <td>601</td> <td>CAAAATGCTTTTCATATTCTTCGCAATGGGTTTACTAGAGAGAAGCAACAG-TTCAAAA</td> <td>659</td>	601	CAAAATGCTTTTCATATTCTTCGCAATGGGTTTACTAGAGAGAAGCAACAG-TTCAAAA	659
Qy <td>3489</td> <td>gctctgaagaagaagtaacattgacttttcatataaggcttcaagattgggaagtcca</td> <td>3548</td>	3489	gctctgaagaagaagtaacattgacttttcatataaggcttcaagattgggaagtcca	3548
Db <td>660</td> <td>GCTCCTCAGAGAAGCAATTAACATTTGACTTTTATCATTAAGGCTTTCAAGATTGGGAAGTTCA</td> <td>719</td>	660	GCTCCTCAGAGAAGCAATTAACATTTGACTTTTATCATTAAGGCTTTCAAGATTGGGAAGTTCA	719
Qy <td>3549</td> <td>gccatgaatacaaatgcttttgggaaaaactcaaaagaaat</td> <td>3589</td>	3549	gccatgaatacaaatgcttttgggaaaaactcaaaagaaat	3589
Db <td>720</td> <td>G-CATGATATACAA---TGCTTTGGAAAAAUAAGGATT</td> <td>756</td>	720	G-CATGATATACAA---TGCTTTGGAAAAAUAAGGATT	756
RESULT	3		
AI361043/C			
LOCUS	AI361043	641 bp	mRNA
DEFINITION	qy03f11.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:2010957 3'	EST	15-FEB-1999
	Similar to TR:P91133 P91133 SIMILAR TOS. CEREBISIAE		
	UBIQUITIN-PROTEIN LIGASE E3 COMPONENT SP:P19812. ; mRNA sequence.		
ACCESSION	AI361043		
VERSION	AI361043.1	GI:4112664	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 641)		
AUTHORS	NCI/NIH/NS-CGAP	http://www.ncbi.nlm.nih.gov/ncicgap.	
TITLE	National Cancer Institute / National Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project (CGAP/BTGP), Tumor Gene Index		
JOURNAL	Unpublished (1998)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D. cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.llnl.gov/bbrp/image/image.html Insert Length: 1083 Std Error: 0.00 Seq primer: -400P from gibco High quality sequence stop: 453. Location/Qualifiers 1..641 /organism="Homo sapiens"		
FEATURES	source		

BASE COUNT		174	a	155	c	135	g	175	t	2 others	
ORIGIN											
/db_xref="taxon:9606" /clone="IMAGE:2010957" /clone_lib="NCI_CGAP_Brn23" /tissue_type="glioblastoma (pooled)" /lab_host="DH10B" /note="organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTCAAGTGGAGCGCGCATATCTTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized, and was constructed by Bento Soares and M.Fatima Bonaldo."											
Query	Match	10.1%	Score	637.4	DB	10	Length	641			
Best	Local Similarity	99.5%	Pred.	No. 2e-138							
Matches	638	Conservative	0	Mismatches	3	Indels	0	Gaps	0		
Qy	5457	cctgagaaaaaagatagtttgaagagcttccctgaactatagctgcctctgaatcaa	5516								
Db	641	CCTGAAAAAGAAATAGTTTGTATAGAGCTTCTGATGACTATAGCTGCCTNTGAATCCA	582								
Qy	5517	gcttctcattcaggtgccacgctctgcagatcagcgaagacactctgtcctctgc	5576								
Db	581	GCTTCTCATNTCAGGTGCCACGGTCTGCAGATGATGAGCGAAGACTCTGTCTCTCTGC	522								
Qy	5577	cttttctgvgggtataactatgttctcagaacatttctgctccaggaaaattgtgaacggg	5636								
Db	521	CTTTTCTGTGGGCTATACTATGTCTCAGAAACATTGTGTCGCCAGGAAATGTGAACGGG	462								
Qy	5637	gaagaggtgagcttgcatctttcaagcactcactcgtggagcggagctgcatcttc	5696								
Db	461	GAAGAGGTTGAGGCTTGCATTTTTTCACGCACCTTCACCTGTGGAGCGGAGTCTGCATTTTC	402								
Qy	5697	ctaaaaatcagaaatgccgagtgctctgttgaagtaaacgacagagctgtgcctat	5756								
Db	401	CTAAAAATCAGAGAAATGCCGAGTGGTCTGTGTGAAGTAAAGCCAGAGGCTGTGCCTAT	342								
Qy	5757	ccagctcctacttgatgaatatggagaacacacccctgacctgaagagggggcaacccc	5816								
Db	341	CCAGTCTCTTACTTGGATGAATATGGAGAAACAGACCTGCGCTGAAGAGGGGCAACCCC	282								
Qy	5817	cttcatttatctcgtgagcgttatcggaagctccatttgcctggcaacaacactgcatt	5876								
Db	281	CTTCATTATCTCGTGAAGCGGTATCGGAAGCTTCCATTGTTGTTGGTGGCAACACACTGCATT	222								
Qy	5877	atagaagagattctagagccaaagagaactaatcagatgttatttgattcaactgcag	5936								
Db	221	ATAGAAGAGATTGCTTAGGAGCCAGAGACTAATCAGATGTTATTTGGATTCACTGGCAG	162								
Qy	5937	ttactgtgagctccaaactctgcctcaagacaatcacaaatgacacagtagtaaaagctg	5996								
Db	161	TTACTGTGAGCTCCAACCTCTGCCCAACACAAATCACAATCAGCAGCTAGTAAGGCTG	102								
Qy	5997	attcaaaaattatggaaaacttctgagggctgggaagtattgagaggtcttttgcctca	6056								
Db	101	ATTCAAAATTTATGGAAAACTTTCTGAGGCTGGGAAAGATTATGGAGGCTCTTTTGCTCCA	42								
Qy	6057	tgtccaggttcaactacatcaataaaaaattttcctaattgga	6097								
Db	41	TGTCAGGTTTCACTTACATCAATAAAAAATTTCTTTAATGGA	1								
RESULT	4										
LOCUS	BC862813	797 bp	mRNA	EST	29-MAY-2001						
DEFINITION	602799074P1 NIH_CGAP_Mam4	Mus musculus	cDNA clone	IMAGE:4934370	5'						
ACCESSION	BG862813	mRNA sequence.									

56062813	BG862813	797 bp	mRNA	EST	29-MAY-2001
LOCUS	60279074F1				
DEFINITION	NIH_CGAP_Mam4 Mus musculus cDNA clone IMAGE:4934370 5',				
	mRNA sequence.				
ACCESSION	BG862813				

Qy	5591	tatactatgtttctcagaacatttctgtccagagaaattgtgaacggggaagaggtlgtgagc	5655
Db	422	CATCCTGTGTCTCAAGAACATCTGTTCGCCAAGAAATAGTGAATGGGGAAGAGGTTGGAGC	481
Qy	5651	ttgcatttttcaacgcacttctactctgtggagccgagctctgcatttctctaaaaatacagaga	5710
Db	482	GTGCGTTTTTCATGCGCTTCATGTGGTGCTGGAGTCTGCATTTTCTCTAAAAATCCGAGA	541
Qy	5711	atgccgagtggtccttggttgaaggttaaagccagagagctgtgcctatccagactccttactt	5770
Db	542	ATGCAGGGTGGTTCCTGGTGAAGAAAGCCAGAGGCTCTGCCTACCCAGCCCTTACTTT	601
Qy	5771	ggatgaatatggagaaaaacagaccctggcctgaagggggcaaccccttcatttatctctcg	5830
Db	602	GGATGAATATGGAGAAACAGACCCAGGGCTAAAGAGAGAGAAACCCACTTCATTTATCTCG	661
Qy	5831	tgagcgggtatcggaagctccatttgctc-tggcaacaacactgcattatagaagaattg	5889
Db	662	GGAGCGGTATCGGAAGCTGCATTTGGTCTTGGCAAAAGGACTTGGATTATAGAAGAATG	721
Qy	5890	ctaggagccaagagagactaatcagatgttatttggattcaactggcagtgacttgagcttc	5949
Db	722	GTGCGAGCGGGGAGACTAATCAGATGCTATTATTTGGACTTAACCTGGGAGTTACTCTGAGATC	781
Qy	5950	caactctgcctccaaga	5965
Db	782	CAGTTGTCTCAAGAGA	797
RESULT 5			
BI086469			
LOCUS	602849734F1	NIH_MGC_10	Homo sapiens cDNA clone IMAGE:4991193 5'
DEFINITION	BI086469	782 bp	mRNA sequence.
ACCESSION	BI086469	EST	20-JUN-2001
VERSION	BI086469.1	GI:14504799	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostom		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 782)		
JOURNAL	NIH-MGC http://mgi.nci.nih.gov/ .		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: Incyte Genomics, Inc. DNA Sequencing by: Incyte Genomics, Inc. Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM11008 row: d column: 10 High quality sequence stop: 715. Location/Qualifiers 1..782 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:4991193" /clone_lib="NIH_MGC_10" /cell_line="MGC36" /lab_host="DH10B" /note="Organ: cervix; Vector: pCMV-SPORT6; Site 1: Not Site 2: SalI; Cloned unidirectionally. Primer: Oligo Average insert size 1.5 kb. Library prepared by Life Technologies."		
BASE COUNT	184 a	194 c	163 g 241 t
ORIGIN			

```
Query Match          9.1%; Score 573.2; DB 11; Length 782;
Best Local Similarity 92.5%; Pred. No. 2.2e-123;
Matches 712; Conservative 0; Mismatches 48; Indels 10; Gaps 10;

QY 4863 aaatcagaagatacacaccatgctctgtctatagatctgtttcatgttttgggtgact 4922
|||||
Db 4 AAATCAGAGATACACCATG-CTTCTGCTATAGATCTGTTACATG-TTTGGTGGATGCT 61
|||||

QY 4923 gtgttagcatcccatctctgttattggatgacccctgttgaactctgcagcctcttcagtt 4982
|||||
Db 62 GTGTTAGCA-TCCCATCTCTGTA-TGGGATGACCCCTGTTGATCTGCAGCCTCTTCAGTT 119
|||||

QY 4983 agttctctatataacacacttattctctccattgtatccatgacacacatgcttcag 5042
|||||
Db 120 AGTTCTCTCTATAACCACTTATCTCTTCCATTTGATCACCATGGCACACATGCTTCAG 179
|||||

QY 5043 atactactacagtagacacagcctacccctgtctcaggttcaagaagacagtagaag 5102
|||||
Db 180 ATACTACTTACAGTAGACACAGCGCTACCCCTTGCTCAGGTTCAAGAAGACAGTAGAAG 239
|||||

QY 5103 gctcattccgcatctcttcttgcagaaattctcctaatacagaagtggtcccatggg 5162
|||||
Db 240 GCTCATTCGCACTCTCTCTTCGCAAGAAATTTCTCAATATACAAAGTGGCTCCATGGG 299
|||||

QY 5163 tgtgatattcctggctggtgattgttgggtctcactgaagaatgg-catacacccttatct 5221
|||||
Db 300 TGTGATATTCCTGGCTGCTGATTTGTTGGGTCTCAGTGAAGATGGCCATCACCCCTTATCT 359
|||||

QY 5222 tgcgtgctgcattggttttccacattattcttgggttaactcgcctgcctgagaaactgca 5281
|||||
Db 360 TCCTGTCTGCTGATTTGTTTTCACATATTTACTTGGGGTAACTCCGCTGAGGAACCTGCA 419
|||||

QY 5282 taccattctcagaagagagatgaca-gtgcactctgtagctatctatttaccataaa 5340
|||||
Db 420 TACCAATCTCGAAGAGAGAGACAGTGCACCTCTGAGCTATCTATCTTTTACCTACAA 479
|||||

QY 5341 attgttctctctctccaggaaattgggatactgtgagggccctgtgtccagagtggt 5400
|||||
Db 480 ATTTGTTCTCTCTTCCAGGAATATTTGGGATACCTGTAAGGCCCTTTTTCAGAGGTGCT 539
|||||

QY 5401 gtgca-gatcctgctctactaa-acgtgttgaaagcaaaaaaacacccgtggtea-ggtacc 5457
|||||
Db 540 GTGAGGATCTCGCTTACTACACTGTTTGAACGCAAAACACACGTGTCAGGTACCC 599
|||||

QY 5458 ctgaaaaaagaattgtttagagcttctgactgactatagctgcctcctgaaatca-a 5516
|||||
Db 600 CTAGACACAGACTAGTTTGTAGAGCTTCCCTGATGAATATAGCTGCTCTGCTGAATCACA 659
|||||

QY 5517 gcttctcattcagtgccacggctctgcagatgatgagcgaagacatcctgtcctctgc 5576
|||||
Db 660 GCTTCTCATCTCAGTGTCCCGGCTCTGCAGATGATTGACCCAAAGCATCTCTCTCCGC 719
|||||

QY 5577 cttttctgtgggctatactattgttctcagaacattgtgtgcacggaat 5626
|||||
Db 720 TCTTCTGTGGGCTATAACTATGCTCCGAACCACTTTGCTGCAGAAAT 769
|||||

RESULT 6
BF063405/c
LOCUS
DEFINITION
  BF063405 565 bp mRNA EST 16-OCT-2000
  7h89d06.x1 NCI-CCAP.Col16 Homo sapiens cDNA clone IMAGE:3323147 3'
  similar to TR:070481 070481 UBIQUITIN-PROTEIN LIGASE E3 COMPONEN
  N-RECOGNIN ; mRNA sequence.
ACCESSION
  BF063405
VERSION
  BF063405.1 GI:10822315
KEYWORDS
  EST.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 565)
  NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
  National Cancer Institute, Cancer Genome Anatomy Project (CCAP),
  TITLE
```

JOURNAL
COMMENT

Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck, M.D.,
Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov

Seq primer: -400P from Gibco
High quality sequence stop: 489.

FEATURES
Source

1..565
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3323147"
/clone_lib="NCI-CCAP.Col16"
/tissue_type="colon tumor, RER+"
/lab_host="DH10B"
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site_1: Not 1; Site_2: Eco RI;
Plasmid DNA from the normalized library NCI-CCAP.Col16 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneIDs 1057416-1061255, and 1144584-1145351).
Subtraction by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 152 a 138 c 115 g 160 t
ORIGIN

Query Match 8.8%; Score 557; DB 11; Length 565;
Best Local Similarity 100.0%; Pred. No. 1.3e-119;
Matches 557; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5541 tctgcagatgtagcgaagcattcctctgctctgctcttctctgtggtggtatactatgt 5600
|||||
Db 565 TCTGTCAGATGATGAGCGAAAGCATCTCTCTCTGCTCTGCTCTGCTGCTGCTGCTGCTGCT 506
|||||

QY 5601 tctcagaacatttgcgcaggaattgtgaacgggaaaggttgagcttgcattttt 5660
|||||
Db 505 TCTCAGAACATTTGCTGCCAGGAATTTGTGAACGGGGAAGGTTGGAGCTTGCATTTT 446
|||||

QY 5661 cagcacttcactgtgagcggagctgctgcatcttctctctctctctctctctctctctct 5720
|||||
Db 445 CAGCGACTTCACCTGTCGAGCGGAGTCTGCATTTTCTTAAATAATCAGAGATGCCGAGTG 386
|||||

QY 5721 gtcctgttggaagttaagccagaggctgtgctctatccatgcctccttacttggatgaatat 5780
|||||
Db 385 GTCTGTGTTGAAGGTAAAGCCAGAGGCTGTGCCTATCCAGCTCCCTTACTTGGATGAATAT 326
|||||

QY 5781 ggagaacagaccctggcctgaagaggggcaaccccttcaatttctcgtgagcgggtat 5840
|||||
Db 325 GGAGAACAGACCTGGCCCTGAAGAGGGGCAACCCCTTCATTATCTGTCGTGAGCGGTAT 266
|||||

QY 5841 cggaactcctcattgttctgtgcaacaacactgcattatagaagagattgctaggagccaa 5900
|||||
Db 265 CGGAAGCTCCATTTGGTCTGCGCAACAACACTGCATATTAGAGAGATTGCTAGGAGCCAA 206
|||||

QY 5901 gagactaatcagatgttatttggattcaactgagcagtgcttactgtgagctcccaactctgct 5960
|||||
Db 205 GAGACTAATCAGATGTTATTTGGATTCAACTGCGAGTTACTGTGAGCTCCAACCTCGCT 146
|||||

QY 5961 caagacaatacaaaatgacacagtagtaaaaggctgattcaaaattatggaaaactttct 6020
|||||
Db 145 CAAGACAATAACAAATGACACAGTAGTAAGAGCGGTGATTCAAAATTTATGGAATACTTCT 86
|||||

QY 6021 gagggctgggaaagtatttgagggtcttttggctccatgctccaggttccacttaccatcaata 6080
|||||


```

Db 85 GAGGCTGGGAAAGTAATTGGAGGCTTTTGTCTCCATGTCAGGTTACATCAATA 26
Qy 6081 aaatatttcttaaggga 6097
Db 25 AAATATTCTTTAATGGA 9

RESULT 7
BE589438 606 bp mRNA EST 28-AUG-2000
LOCUS 195602 BARC 5BOV Bos taurus cDNA 5', mRNA sequence.
DEFINITION BE589438
ACCESSION BE589438
VERSION BE589438.1 GI:9842477
KEYWORDS EST.
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 606)
Source: Sonstegard, T.S., Capuco, A.V., Van Tassel, C.P., Ashwell, M.S. and
Wells, K.D.
Mapping of Expressed Sequence Tags from a normalized bovine mammary
gland cDNA library
Unpublished (2000)
Contact: Sonstegard TS
USDA, ARS, Beltsville Agricultural Research Center
Bldg. 200 Rm 2A, Beltsville, MD 20705, USA
Tel: 301 504 8416
Fax: 301 504 8414
Email: tads@psi.barc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -mismatch 12 options.
PCR Primers
FORWARD: AGCAACAGCTATGACCAT
BACKWARD: GTTTCACAGTCACGAG
Plate: 120 row: N column: 11
Seq primer: ATTAGGTGACACTATAG.
FEATURES
source
Location/Qualifiers
1..606
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="BARC 5BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled mRNA isolated from mammary
tissues at eight physiological, developmental, and disease
states."
BASE COUNT 144 a 146 c 113 g 203 t
ORIGIN

Query Match 8.1%; Score 512.2; DB 10; Length 606;
Best Local Similarity 90.4%; Pred. No. 4.2e-109;
Matches 547; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

Qy 4761 aaagcattaatgcagttgcagttgcagagagattaccctgtctctcaagtcctgatacag 4820
Db 1 AAAGCATTAATGCAAGTTGTCATTGCAACAAAGGATTACCTGTTCTCAGGCTCGATACAG 60

Qy 4821 aaacatcgttgcgtctctatcagttgtcttcctaaacataaaatcagaagatacacca 4880
Db 61 AAACATCTGAFTTCCTTCATACAGTTGTTCTTCCTAACCTGCAATCAGAGCTACACCT 120

Qy 4881 tgcctctgtctatagatctgtttcattcagttgttggtggtgctgtgttagcattccccatcc 4940
Db 121 TGCCTTCATCCATAGATCATTTTCATGTTTGTGGGTACTGTGTAGCATTCCTCCGTCC 180

Qy 4941 ttgtattggatgaccctgtgtatcgtcagcctctctctcagttagttcttctctataaccac 5000
Db 181 TTATACTGGGATGATGCTGTGTGATCTGCAGCCCTTCACCAAGTTAGCTTCTCGTATAACCA 240

```

```

Qy 5001 ctttatctctccatttgatcacccatgagcacacatgcttcaagatactactacagtagac 5060
Db 241 CTTTATCTCTTCCATTGATCACCATGGCACACATGCTTCAGATACTCTTTACCATAGAC 300

Qy 5061 acagccctacccttgctcaggttcaagaagacagtgaaagaggtcattccgcattctct 5120
Db 301 ACAGACCTCCCTCTGCACAGATACAGGAAGAGAGTGAAGAGGCTCGTCTGCATCTTCT 360

Qy 5121 tctttgcagaaatttctcaatacaaatggtccattgggtggtgatatctctcgtgctg 5180
Db 361 TTCCTGGCAGAAAGTTCTCAGTATACAAAGTGGCTGCATGGGTGGTATTCCTGGCTGG 420

Qy 5181 tatttgggtctcactgaagaatggcacccttatcttcgtgtgctgctgattgttt 5240
Db 421 TATTATGGGTCTCCTTGAAGAATGTCATCCCGTATCTTCGCTGCTGCTGCATCTTCT 480

Qy 5241 ttccactatttactgggttaactccgctgaggaactgcatacaattctgcagaagga 5300
Db 481 TTCCCACTATTATTGGAGTAACCTCCACCTGAAGAATCTGTTACCAATGCTGAGAAGGA 540

Qy 5301 gagtacagtcactgtgtagctatctcttaccacaaatttgcctgctctccag 5360
Db 541 GAATACAATGTACTGTGTAGCTATTATTATCTTACCACAAATTTGTTCTGCTTTCCAC 600

Qy 5361 gaata 5365
Db 601 AAATA 605

RESULT 8
BG219270/c 624 bp mRNA EST 21-APR-2001
LOCUS RST39023 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
DEFINITION BG219270
ACCESSION BG219270
VERSION BG219270.1 GI:13745291
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 624)
AUTHORS Harrington, J.J., Sherf, B., Rundlett, S., Jackson, P.D., Perry, R.,
Cain, S., Dahl, T., Thornton, M., Ramachandran, R., Whittington, J.,
Lerner, L., Krashoc, D., McElligott, K., Clark, S., Mays, R., Smith, E.,
Velo, N., Hess, J., Cothren, K., Lo, K., Offenbacher, J., Danzig, J.,
and Ducar, M.
TITLE Creation of Genome-wide Protein Expression Libraries using Random
Activation of Gene Expression
JOURNAL Nat. Biotechnol. 19 (5), 440 (2001) In press
COMMENT Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scain@athersys.com
High quality sequence stop: 624.
FEATURES
source
Location/Qualifiers
1..624
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="HT1080"
/note="See 'Creation of Genome-wide Protein Expression
Libraries using Random Activation of Gene Expression',
Nature Biotechnology, in press. Note that even though the
cell type indicated is HT1080, since a random activation
method was used, these sequence tags are not necessarily
expressed in HT1080 under normal circumstances."
BASE COUNT 129 a 151 c 136 g 206 t 2 others
ORIGIN

```

/dev_stage="adult"
/lab_host="DHI0B"
/note="Organ: uterus; Vector: pT73-Pac; Site_1: Not I;
Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5',
AATGGAAGAAATTCGGCCCGCTTTTGTCTTTTTTTTTTTT 3'] ,
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by M. Fatima Bonaldo."

BASE COUNT 125 a 96 c 86 g 174 t

ORIGIN

Query Match 7.6%; Score 479.8; DB 11; Length 624;
Best Local Similarity 98.2%; Pred. No. 1.6e-101;
Matches 484; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 3723 attactcatgataaaagcagaacgaagaaaagaaagctgaaagtctaggctacat 3782
|||||
Db 624 ATTACTCATGATAAAGAAAGCAGAACCAAGAAAAGAAAGAAAGCTGNAGCTNTAGTTACAT 565
|||||
QY 3783 cgccagaagatcatggctcagatgtctgccttacagaaaaacttcattgaaactcataaa 3842
|||||
Db 564 CGCCAGAAGATCATGCCTCAGATGCTGCTCTACAGAAAAAATTCTATTGAAACTCATAAA 505
|||||
QY 3843 ctcatgtatgacaatacacatcaagaatcgctgggaagaagattccattatgaggaagag 3902
|||||
Db 504 CTCATGTATGACAATACATCAGAAATGCCCTGGGAAAAGAAAGATTCCATTATGGAGGAAGAG 445
|||||
QY 3903 agaccgccagctcagtaactactactctaagaattgcttggttgcctaaacgggtccatct 3962
|||||
Db 444 AGACCCAGCAGCTAGTGACTACTCTAGAATGCTTTGGGTGCTTAACGGGGTCCATCT 385
|||||
QY 3963 gt tactgaaaaaggggtgctgcagctgcattcctttgccagaagaacagaggtgaaaaata 4022
|||||
Db 384 GTTACTGAAAAGGAGGTGCTGAGCTGCATCTTGTGCCNAGAGAACAGAGGTGAAAAATA 325
|||||
QY 4023 gaaaaaatgccaatgatattatggcctgtgtocagaaatctactgccttaaccaccagac 4082
|||||
Db 324 GAAAATATGCCATGTATTATTCGGCTGTGTCACAGAAATCTACTGCCTTAACCCAGCAC 265
|||||
QY 4083 aggggaaaccatagaacctcagagaagcccctagaccactttteatgatccagac 4142
|||||
Db 264 AGGGGAAACCCTAGAACTCTCAGAGAGAGCCCTAGACCCCCTTTTCATGATGCCAGAC 205
|||||
QY 4143 ttggcatatggaacttatcagagaagctgtggtcatgtaatgcacgcagctgtcgtggcag 4202
|||||
Db 204 TTGGCATATGGAACCTTATACAGGAACCTGTGTCATGTAATGCACGCAGTGTGTCGGCAG 145
|||||
QY 4203 aagtatttgaag 4215
||| | |||
Db 144 AACTCCTCCGGAG 132

RESULT 9

AI192195/c

LOCUS

DEFINITION

q92e08.x1 Soares_pregnant_uterus_NbhPU Homo sapiens cDNA clone IMAGE:1721702 3' similar to TR:O15057 O15057 KIA0349 ;, mRNA sequence.

ACCESSION

AI192195

VERSION

AI192195.1 GI:3743404

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 {bases 1 to 481}

AUTHORS

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE

National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL

Unpublished (1997)

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (infoimage.llnl.gov) for further information.
Insert Length: 563 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 351.
Location/Qualifiers
1..481
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1721702"
/clone_lib="Soares_pregnant_uterus_NbhPU"
/sex="female"

FEATURES

source

/dev_stage="adult"
/lab_host="DHI0B"
/note="Organ: uterus; Vector: pT73-Pac; Site_1: Not I;
Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5',
AATGGAAGAAATTCGGCCCGCTTTTGTCTTTTTTTTTTTT 3'] ,
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by M. Fatima Bonaldo."

BASE COUNT 125 a 96 c 86 g 174 t

ORIGIN

Query Match 7.6%; Score 477.8; DB 10; Length 481;
Best Local Similarity 99.6%; Pred. No. 4.6e-101;
Matches 479; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2778 ctccagattggtgcatctttaatggatcccccaataagttcttgttactgttacttcagag 2837
|||||
Db 481 CTTTCAGATTGGTGTCATCTTTTAAATGGATCCAATAAGTTCTTGTACTTGCTTTCAGAGG 422
|||||
QY 2838 tatgaacttgcgcagggcttttaacaagaccatatctacaaaagaccagatttgattaaa 2897
|||||
Db 421 TATGAACCTTGGCCAGGCTTTTAAACAGACCATATCTACAAAAGCCCGAGGATTGATTAAA 362
|||||
QY 2898 caataataacactaataagaagaatgcttcaggctcctctatctatttgtggtagcgt 2957
|||||
Db 361 CAATATAATACACTAATAGAAGAAATGCTTCAGGTCCCTCATCTATATTGTGGTGAGCGT 302
|||||
QY 2958 tatgtacctggagtgggaaatgtgaccaagaagaggtcacaaatgagagaaacattcac 3017
|||||
Db 301 TATGTACCTGGAGTGGGAAATGTGACCAAAGAGAGGTACAAATGAGAGAAATCATTCAC 242
|||||
QY 3018 ttgcttgcattgaaccocatgccacacagtgccattgcccataaacatttacctgagaatgaa 3077
|||||
Db 241 TTGCTTTGCAATTGAACCCATGCCACACAGTGCCATTGCCAAAAATTTTACCTCAGAAATGAA 182
|||||
QY 3078 aataatgaactggcttagaagaatgtcataaacaagaagtgccacatttaagaaccagct 3137
|||||
Db 181 AATAATGAACCTGGCTTAGAGATGTATGAATAAAGAGTGGCCACATTTTAGAAACCCAGGT 122
|||||
QY 3138 gtatcaggccatggagtttatgaactaaaagatgaatcactgaagacttcaatatgtac 3197
|||||
Db 121 GTATCAGGCCATGGAGTTTATGAACATAAAGATGAATCACTGAAAGACITCAATATGTAC 62
|||||
QY 3198 ttttatcattctccaaacccagcatagcaggctgaacatatgcagaagaaaaaggaga 3257
|||||
Db 61 TTTTATCATTTACTCCAAAAACCCAGCATAGCAAGGCTTGAACATATGCAAGAAAAAGAGAG 2

QY 3258 a 3258

Db 1 A 1

RESULT 10

AW971391

LOCUS

AW971391

DEFINITION

EST383480 MAGE resequences, MAGL Homo sapiens cDNA, mRNA sequence.

ACCESSION

AW971391

VERSION

AW971391.1 GI:8161236

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 {bases 1 to 583}

AUTHORS

Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspar,R., Gay,C., Holt,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeaman,T.J. and Quackenbush,J.

TITLE

Assessment of gene expression patterns in a model of colon tumor metastasis using a 19,200 element cDNA microarray

JOURNAL
COMMENT

Unpublished (2000)
Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@tigr.org
Plate: 289

Seq primer: Forward.

Location/Qualifiers

1. .583
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:resquences, MAGL"
/note="Vector: pBluescriptSKm"

FEATURES

source

171 a 126 c 137 g 149 t
Location/Qualifiers

/organism="Homo sapiens"
/db_xref="GDB:5929590"
/db_xref="taxon:9606"
/clone_lib="IMAGE:742663"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"

Query Match 7.5%; Score 473.4; DB 10; Length 583;

Best Local Similarity 99.8%; Pred. No. 5.1e-100;

Matches 474; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5678 agccgagctgcatttctctaaatacagagaatccgagtgctcgtgtaagataa 5737

Db 105 AGCCGAGGCTGCATTCTCTAAATACAGAGATCCGAGTGTCTGTTGAAGTAA 164

QY 5738 agccgaggtgtgcctatccagctccttacttgatgaatggagaacagaccctgg 5797

Db 165 AGCCAGAGGCTGCTATCCAGCTCCTTACTTGGATGATATGGAGAACAGACCTGG 224

QY 5798 cctgaagaggggcaacccctcattatcctgtagcggtatcgagaagctccattgggt 5857

Db 225 CTGGAAGAGGGCAACCCCTTCATTATCTCGTAGCGGTATCGGAAGCTCCATTGGT 284

QY 5858 ctggcaacaacactgatataagaagattgtagagagccagagacatacatcagattt 5917

Db 285 CTGGCAACAACACTGCTATTATAGAAGAGATTGCTAGAGCCAGAGACTAATCAGATGT 344

QY 5918 atttgattcaactggcagtgactgagctccaactctcctcaagacatacacaaatg 5977

Db 345 ATTGTGATTCACTGGCAGTACTGTGAGCTCCAACCTCGCTCAGACAAATCACAATG 404

QY 5978 acgacagtagtaagagctgattcaaaattatggaaaacttctcagggctgggaaagtat 6037

Db 405 ACACAGTAGTAAAGGCTGATTCAAAATTATGGAATAATTCTGAGGGCTGGGAAAGTAT 464

QY 6038 tggaggtgtttgctccatgctcaggttcaactacatacaataatttcttaattgga 6097

Db 465 TGGAGGGTCTTTTGCCTCATGTCCAGGTTCACATTACATCAATAAATAATTCTTAATGGA 524

QY 6098 gtattgtttcaattagcaacatatgtcttcacaggaagaaagacatagatcaa 6152

Db 525 GTATTGCTTCATTAGCAACATATGCTTCACAGGAAAGGACATAGATCCA 579

RESULT 11

AA401319

LOCUS

AA401319 478 bp mRNA EST 09-NOV-1997

DEFINITION

Zu63d04.r1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:742663

5', mRNA sequence.

ACCESSION

AA401319

VERSION

AA401319.1

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 478)

Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,

Krisman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin

J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,

White, Y., Wylie, T., Waterston, R. and Wilson, R.

TITLE
JOURNAL
COMMENT

WashU-NCI human EST Project
Unpublished (1997)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

This clone is available royalty-free through LINL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 1003 Std Error: 0.00

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 458.

FEATURES

source

1. .478
Location/Qualifiers

/organism="Homo sapiens"
/db_xref="GDB:5929590"
/db_xref="taxon:9606"
/clone_lib="IMAGE:742663"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"

/note="Vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech Laboratories
, Inc., and primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGAAGTGGAGCGGCCCAATTTTTTTTTTTT 3'].

Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified p7T3 vector. Library

went through one round of normalization to Cot5, and was

constructed by Bento Soares and M. Fatima Bonaudo. "

BASE COUNT 114 a 109 c , 114 g 141 t

ORIGIN

Query Match 7.5%; Score 471.6; DB 10; Length 478;

Best Local Similarity 99.2%; Pred. No. 1.3e-99;

Matches 474; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 5173 ctgctgtatttgggtctcactgaagaatggcatcaccccttatcttcgtgctg 5232

Db 1 CTGGCTGTATTGTGGGTCTCCTGGAAGATGGCATCACCCCTTAICTTCGCTGTCG 60

QY 5233 cattgttttccactatttacttgggttaactccgctcctgagaaactgcatacaattctg 5292

Db 61 CATTTGTTTTCCACTATTACTTGGGGTAACCTCCGCTCAGGAACCTGCATCAATCTG 120

QY 5293 cagaagagagtagcagtgcaactctgtagctatctatttaccatacaaatgttctctgc 5352

Db 121 CAGAAGGAGAGTACAGTGCACCTCTAGCTATCTATCTTTACCTGCAAAATTTGTTCTG 180

QY 5353 tcttccaggaatattgggatactgaagcccttgcctcagaggtggtgtcagatccctg 5412

Db 181 TCTTCCAGGAATATGGGATATCTGAAGCCCTTCTCAGAGGTGCTGTCAGATCTCTG 240

QY 5413 ccttactaaactgtttgaagcaaaaaacacccgtggtcaggtaccctagaaaaaata 5472

Db 241 CCTTACTAAACTGTTGAAGCAAAAAACACCCGTGTCAGGTACCTAGAAAAAATA 300

QY 5473 gttgtagagacttctgtagtactatagctgctcctgaatcaagcttctcattcaggt 5532

Db 301 GTTTGATAGAGCTTCTCTGATGCTAGTGCCTCCTGAATCAAGCTTCTCATTTTCAGGT 360

QY 5533 gccacaggtctgcagatgatgagcgaaagcatcctgtcctctgcttcttctgtgggcta 5592

Db 361 GCCCAGCGTCTGCAGATGATGAGCGAAAAACAGTCTCTCTCTGCTCTCTGCTGGGCTA 420

QY 5593 tactatgtcttcagaacacatttgcgtccaggaattgtgaacggggaagaggttgagc 5650

Db 421 TACTATGTTCTCAGAACATTTTGTCTCCAGGAAATTTGTGAACGGGGAAGAGTTGGAGC 478

```

RESULT 12
AW291190/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 522)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. cDNA library preparation: M.B. Soares Lab Clone distribution:
NCI-CGAP clone distribution information can be found through the
I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Seq primer: M13 Forward
POLYA=Yes.

FEATURES
source
Location/Qualifiers
1..522
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2723895"
/clone_lib="NCI-CGAP_Sub4"
/clone_host="DH10B (Life Technologies)"
/note="Vector: pT7SD-Pac (Pharmacia) with a modified
polylinker; Site.1: Not I; Site.2: Eco RI; The
NCI-CGAP_Sub4 library is a subtracted library derived from
the NCI-CGAP_Sub2 library which is a subtracted library
derived from the NCI-CGAP_Sub1 library, which is a
subtracted library derived from BI. BI constitutes a
mixture of 21 normalized or subtracted NCI-CGAP
libraries: NCI-CGAP_Co4, NCI-CGAP_Pr22, NCI-CGAP_Pr28,
NCI-CGAP_Co10, NCI-CGAP_Co16, NCI-CGAP_Kid5,
NCI-CGAP_Kid12, NCI-CGAP_Kid3, NCI-CGAP_Kid11,
NCI-CGAP_Lym2, NCI-CGAP_Br2, NCI-CGAP_Co8, NCI-CGAP_CLL1,
NCI-CGAP_Le12, NCI-CGAP_Brn23, NCI-CGAP_Lu5,
NCI-CGAP_Lu24, NCI-CGAP_Lu19, NCI-CGAP_GC4, NCI-CGAP_GC6,
NCI-CGAP_Brn25. These 21 libraries were pooled and a
single-stranded DNA preparation of the resulting mixture
was used as a tracer in a subtractive hybridization with
a driver whose composition is detailed below:
NCI-CGAP_Kid3 pool 1 : LLAM 3334-3337, 3682-3683,
3798-3803 (IMAGE Clonids 1322376-1323911,
1456008-1456775, 1500552-1502855) NCI-CGAP_Kid5 pool 1 :
LLAM 3338-3342, 3722-3725, 3776-3778 (IMAGE Clonids
1323912-1325831, 1471368-1472903, 1492104-1493255)
NCI-CGAP_Lu5 pool 1 : LLAM 3375-3582, 3851-3854 (IMAGE
Clonids 1414920-1417991, 1520904-1522439) NCI-CGAP_GC4
pool 1 : LLAM 3164-3167, 3716-3720, 3733-3735 (IMAGE
Clonids 1257096-1258631, 1459064-1470983, 1475592-1476743
) NCI-CGAP_Pr22 pool 1 : LLAM 2457-2459, 2758-2759,
3062-3068 (IMAGE Clonids 985608-986759, 1101192-1101959,
1217928-1220615) NCI-CGAP_Co10 pool 1 : LLAM 2644-2653,
2871-2872 (IMAGE Clonids 1057416-1061255, 1144584-1145351
) Subtraction was performed as previously described
[Bonaldo, Lennon & Soares (1996): Normalization and
Subtraction: Two Approaches To Facilitate Gene Discovery.
Genome Research 6, 791-806.]
TAG_LIB=NCI-CGAP_Co4
TAG_TISSUE=colon
TAG_SEQ=CTTCG"
137 a 125 c 100 g 160 t
BASE COUNT

```

ORIGIN

```

Query Match          7.2%; Score 454; DB 10; Length 522;
Best Local Similarity 98.0%; Pred. No. 1.7e-95;
Matches 492; Conservative 0; Mismatches 5; Indels 5; Gaps 3;

Qy 5596 tatgtctcagaacattgtgtccaggaattgtgaacggggaagaggttggagcttgca 5655
      |||||
Db 513  TATGTTCTCAGAACATTTGCTGCCAGGAAA-TGTGAACGGGGAAGAGTTGGAG---TTCA 458
      |||||
Qy 5656 ttttcacgcacttcactgtgtgagccggagtgctgcattttctctaaataacagagaatgcc 5715
      |||||
Db 457  TTTTTCACGCACCTTCACTGTGGAGCGGAGTCTGCATTTTTCCTAAAAATCAGAGAATGCC 398
      |||||
Qy 5716 gagtgctcctggttgaaggttaaacccagaggtgtgctctatccagctctcttacttgatg 5775
      |||||
Db 397  GAGTGTCTCTGGTTGAAGGTAAGCCAGAGGCTGTGCCTTATCCAGCTCCTTACTTGGATG 338
      |||||
Qy 5776 aatatggagaacacagaccctggccctgaagaggggcaaccccttctattctcgtgagc 5835
      |||||
Db 337  AATATGGAGAAACACACCCTGGCCTGAAGAGGGGCAACCCCTTCATTTATCTCTGTTGAGC 278
      |||||
Qy 5836 ggtatcggaaagctccatttggctctgtgcaaacacactgcattatagaagagattgctagga 5895
      |||||
Db 277  GGTATCGGAAGCTCCATTTGGTCTGSCAACACACTGCAATTATAGAAGAGAGATTGCTAGGA 218
      |||||
Qy 5896 gccaaagactaatcagatgttatttggattcaactggcagttactgtgagctccaaactc 5955
      |||||
Db 217  GCCAAGACAGCAATCAGATGTTATTGGATTCAACTGGCAGTTACTGTGAGCTCCAATC 158
      |||||
Qy 5956 tgcctcagaacaatcacaaatgacacagctagtaagagctgattcaaaaattatggaaaac 6015
      |||||
Db 157  TGCCTCAGACAACATCACAATGACGACAGTAGTAAGGCTGATTCAAAATATTGGAAC 98
      |||||
Qy 6016 ttcttgagggctgggaagattggagggctcttttctccatgtccaggttcacttacat 6075
      |||||
Db 97  TTTCTGAGGGCTGGGAAAGTATTGGA-GTCTTTTCTCCATGTCCAGGTTCACTTACAT 39
      |||||
Qy 6076 caataaataattctctaagga 6097
      |||||
Db 38  CAATAAAATATTCTTAATGGA 17
      |||||

RESULT 13
BG382648          522 bp mRNA          EST          12-MAR-2001
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
1 (bases 1 to 522)
Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
and Keele,J.W.
Design and use of two pooled tissue normalized cDNA libraries for
EST discovery in swine
Unpublished (2000)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smithem@mail.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT

```

BACKWARD: GTTTCGCCAGTCACGACG
Plate: 4 row: H column: 19
Seq primer: ATTTAGTGACACTATAG.
FEATURES
source Location/Qualifiers
1..522

Design and use of two pooled tissue normalized cDNA libraries for EST discovery in swine
Unpublished (2000)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore and -mismatch 12 options.

BASE COUNT	122 a	126 c	100 g	174 t
ORIGIN				

Query Match 7.1%; Score 449; DB 11; Length 522;
Best Local Similarity 91.4%; Pred. No. 2.6e-94;
Matches 476; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

[illegible]

Qy 4979 agttagtcttctctataaacacacctttatctcttccatttgatcaccatggcacacatgct 5038
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 61 AGTTAGCTCTTCCTATAAACCCTTTATCTCTCCATTGATCACCATGGCACATGTT 120

Db 121 TCAGATACTTCTTACCATACACACAGGCCTTCCCTTGACACAGCTCAAGAAGACAGAGTGA 180

Qy 5099 agaggctcatcgcgcattctcttctcttgcagaattctcaatatacaagtggctccat 5158

D6 181 AGAGGCTCATTTCGCATCTTCCTTTCTTGGCAGAAGTTTCTCAATATACAAATAGGCTACGT 240

Db 241 TGGGTGGTATTCTCGGTGGGTATTATGGGTCTCATTTGAAGAAATGGCATCATCCCCTTA 300

Qy 5219 tcttcgctgtgctgcattgttttccactatttacttggggttaactccgcctgaggaact 5278

[illegible]

Db 361 GTTTACCAATTCTGCAGAGAGCGAGTACAGTGCACCTCTGTAGCTATTATCTTTACCCAC 420

Qy 5339 aaattigtctgtctttccaggaaatttgggatactgttaaggcccttgcctccagaggtg 5398

DD 421 AAATTTGGTTCCTGCTTTCCAGAAATAATTGGGATACATTAAGGCCCTCCTCCAGAGGIG 480

O:: 5200 ctctcagcat cctccggttcactcaacatctcttctgcggcgaaaaaa 5420

Db 481 GTGTGCAGACCCCTGCCTTACTAACTGTTTGAGCAAAAA 521

RESULT 14

LOCUS	520 bp	EST	12-MAR-2001
BG382624	BG382624	mRNA	
DEFINITION	298590 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence.		

ACCESSION	BG382624
VERSION	BG382624.1
	GI:13307096

KEYWORDS: LSI, dig
SOURCE:

SOURCE	ORGANISM	Fig.	Sus	scrofa
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24				
25				
26				
27				
28				
29				
30				
31				
32				
33				
34				
35				
36				
37				
38				
39				
40				
41				
42				
43				
44				
45				
46				
47				
48				
49				
50				
51				
52				
53				
54				
55				
56				
57				
58				
59				
60				
61				
62				
63				
64				
65				
66				
67				
68				
69				
70				
71				
72				
73				
74				
75				
76				
77				
78				
79				
80				
81				
82				
83				
84				
85				
86				
87				
88				
89				
90				
91				
92				
93				
94				
95				
96				
97				
98				
99				
100				

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

REFERENCE
AUTHORS
I (cases 1 to 320)
Fahrenkrug, S.-C., Preking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E.,
Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.A., Laegreid, W.W.
and Keele, J.W.

RESULT 15

BE550873/c
LOCUS BE550873 486 bp mRNA EST 06-SEP-2000
DEFINITION UI-M-BH3-ata-a-07-0-UI.r1 NIH_BMAP_M_S4 Mus musculus cDNA clone
UI-M-BH3-ata-a-07-0-UI 5', mRNA sequence.
ACCESSION BE550873
VERSION
KEYWORDS
SOURCE EST.
ORGANISM house mouse.
Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 486)
REFERENCE Bonaldo,M.F., Lennon,G. and Soares,M.B.
AUTHORS Normalization and subtraction: two approaches to facilitate gene
TITLE discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 9704477
COMMENT Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mEST@mail.nih.gov
cDNA Library Preparation: M.B. Soares Lab Clone Distribution:
Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It
should be noted that Bento Soares is generating a small number of
additional specialized non-redundant arrays of BMAP cDNAs whose
availability will be considered under appropriate and limited
collaborative arrangements
Seq primer: M13 Reverse.
Location/Qualifiers
1. .486
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-BH3-ata-a-07-0-UI"
/clone_lib="NIH_BMAP_M_S4"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note=Vector: pR73d-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The
NIH_BMAP_M_S4 library is a subtracted library of a series,
ultimately derived from a mixture of individually tagged
normalized libraries from ten regions of the mouse brain
(cerebellum, brain stems, olfactory bulbs, hypothalamus,
cortex, amygdala, basal ganglia, pineal gland, striatum,
hippocampus) after a series of subtractions to reduce the
representation of cDNAs from which ESTs had already been
generated. The following serially subtracted libraries
were generated in this process: NIH_BMAP_M_S4,
NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, NIH_BMAP_M_S3.1,
NIH_BMAP_M_S2, NIH_BMAP_M_S1. The subtracted library
(NIH_BMAP_M_S4) was constructed as follows: PCR amplified
cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and
NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived
was used as a driver in a hybridization with a pool of
the NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1
libraries in the form of single-stranded circles. The
remaining single-stranded circles (subtracted library)
was purified by hydroxyapatite column chromatography,
converted to double-stranded circles and electroporated
into DH10B bacteria (Life Technologies) to generate the
NIH_BMAP_M_S4 library. This procedure has been previously
described (Bonaldo, Lennon and Soares, Genome Research
6:791-806, 1996).
BASE COUNT 129 a 117 c 95 g 145 t
ORIGIN

QY 2079 cagggttagccagacacaaattgggaagagatatatgcagtaatatgtgacctaaagtat 2138
|||||
Db 486 CAGGGTTATAGCCAGCACAAACTGGGAAGAGTCTACGCAGTATATGTGACCTAAAGTAT 427
QY 2139 atcctgatcagcaaaacccacaatatggacagaagattaaagaatgcagttccttgaaggt 2198
|||||
Db 426 ATCCTGATTAGCAAGCCTGTTCATATGGACAGAACGATTAAAGAGCGCAGTTCTCTGCAAGGT 367
QY 2199 ttctcatcttttttgaagattcttacctgtatcacagggaatggaagaaatccgaagacag 2258
|||||
Db 366 TTCCGGCTCTTTCTGAAGATCTTACCTGTATGCAGGGAATGGGAAGAAATCAGAGACAA 307
QY 2259 gttggcacaacattgaagtggatcctctgtatgggaggtgccattgctatacacagatcaa 2318
|||||
Db 306 GTTGGACAACACATTCAAGTGGACCCCTGACTGGGAGGCTGCCATCGTATACACATGCAA 247
QY 2319 ttgaagaatatcttttactcatgttccaaagagtggtgtgtgtgatgaagaactcttactt 2378
|||||
Db 246 CTAAGCAATATTTTGTCTCATGTTCCTCAAGAGTGGTGTGTTGTGATGAAGATCTCTTACTG 187
QY 2379 gtggcttataaagaatgtcacaaagctgtgatgaggtgcagtagccagtttccatctagt 2438
|||||
Db 186 GTGGCTTATAAAGATGTCAAAAGCTGTAATGAGGTGCAGTACAAATTTTCATGTCCAGT 127
QY 2439 agcaagacagtagtacacaatcgtgtggacatagtttggaaacaaagtcctacagagtatct 2498
|||||
Db 126 ACCAAGACAGTAGTGCATTAATGTGGGTCTAGTCTGGAAACCAATCTCTACAAAGTGTCT 67
QY 2499 gaggatcttgaagcatcacatctgccactctctaggaccttgcgtgtcttcatgtacgt 2558
|||||
Db 66 GAGGACCTTTGTAAGCATACACCTGCCACTCTCTAGAACACTTGTGCTGCTTTCATGTACGT 7
QY 2559 ttaagc 2564
|||||
Db 6 TTAAGC 1

Search completed: March 1, 2002, 14:46:36
Job time: 8041 sec

Query Match 6.5%; Score 410.8; DB 10; Length 486;
Best Local Similarity 90.3%; Pred. No. 2.3e-85;
Matches 439; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 1, 2002, 13:51:46 ; Search time 261.48 Seconds
(without alignments)
20682.299 Million cell updates/sec

Title: US-09-724-126A-1
Perfect score: 6308
Sequence: 1 gccagaattcgcacgagg.....aatttgatttggtgtttt 6308

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues
Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_1101.*
1: /SID2/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SID2/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SID2/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SID2/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SID2/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SID2/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SID2/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SID2/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SID2/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SID2/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SID2/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SID2/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SID2/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SID2/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SID2/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SID2/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SID2/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SID2/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SID2/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SID2/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SID2/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SID2/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4435.4	70.3	6395	20 AAV99308	CDNA encoding mous
2	4435.4	70.3	6395	22 AAC86933	Nucleotide sequenc
3	2546.8	40.4	2550	22 AAH14878	Human CDNA sequenc
4	996.2	15.8	1001	20 AAX03300	Partial CDNA encod
5	996.2	15.8	1001	22 AAC86934	Nucleotide sequenc
6	813.6	12.9	3327	20 AAX35730	CDNA encoding a pr
7	813.6	12.9	3502	20 AAX35731	CDNA encoding a pr
8	656.8	10.4	733	21 AAA02411	Human colon cancer
9	647	10.3	712	22 AAH07621	Human CDNA clone (
10	638.6	10.1	756	21 AAA02327	Human colon cancer
11	246.8	3.9	807	20 AAX39891	Gastric cancer ass

12	241.6	3.8	264	21	AAA00233	Human colon cancer
13	191.6	3.0	800	20	AAV39892	Gastric cancer ass
14	159	2.5	455	20	AAV89674	EST clone DA490.
15	118	1.9	936	22	AAF58252	Oligonucleotide D1
16	118	1.9	936	22	AAF58254	Oligonucleotide D1
17	118	1.9	936	22	AAF58257	Oligonucleotide D1
18	118	1.9	936	22	AAF58259	Oligonucleotide D2
19	118	1.9	936	22	AAF58262	Oligonucleotide D2
20	118	1.9	938	22	AAF58255	Oligonucleotide D1
21	116.6	1.8	308	16	AAT21876	Human gene signatu
22	116.4	1.8	936	22	AAF58252	Oligonucleotide D1
23	116.4	1.8	936	22	AAF58254	Oligonucleotide D1
24	116.4	1.8	936	22	AAF58257	Oligonucleotide D1
25	116.4	1.8	936	22	AAF58259	Oligonucleotide D2
26	116.4	1.8	936	22	AAF58262	Oligonucleotide D2
27	116.4	1.8	938	22	AAF58255	Oligonucleotide D1
28	105.6	1.7	366	21	AAC00938	Human secreted pro
29	105.6	1.7	456	21	AAC00940	Human secreted pro
30	105.6	1.7	1461	21	AAC78096	Human cancer assoc
31	103.2	1.6	605	21	AAC00939	Human secreted pro
32	100.8	1.6	602	21	AAC10733	Human secreted pro
33	100.8	1.6	692	21	AAC10734	Human secreted pro
34	100.8	1.6	842	21	AAC10872	Human secreted pro
35	54.8	0.9	2793	21	AAV70146	Plasmodium falcipa
36	52.4	0.8	244	22	AAF58238	Oligonucleotide D1
37	51.6	0.8	244	22	AAF58238	Oligonucleotide D1
38	49.2	0.8	1727	22	AAH17089	Human CDNA sequenc
39	48.5	0.8	4590	22	AAH24065	Yeast AOD9604-asso
40	46.2	0.7	581	22	AAH09232	Human CDNA clone (
41	45.2	0.7	11722	19	AAV34455	Human MHC class I
42	45	0.7	143068	21	AAF21105	Human low adenosin
43	45	0.7	143068	21	AAF21272	Human low adenosin
44	45	0.7	143068	21	AAA34983	Human adenosine re
45	45	0.7	143068	21	AAA35150	Human adenosine re

ALIGNMENTS

RESULT 1
AAV99308
ID AAV99308 standard; cdna; 6395 BP.
XX *
AC AAV99308;
XX
DT 25-MAR-1999 (first entry)
XX
DE CDNA encoding mouse a ubiquitin-protein ligase, Ubql.
XX
KW Ubiquitin-protein ligase; Ubql; mouse; ubiquitinylation; degradation;
KW N-end rule pathway; stress-related muscle wasting; inhibitor; screen; ss.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT CDS 115..5388
FT /*tag= a
FT /product= Ubql
XX
XX US5861312-A.
XX
XX PD 19-JAN-1999.
XX
XX PF 02-DEC-1997; 97US-0982956.
XX
XX PR 02-DEC-1997; 97US-0982956.
XX
XX PA (CALY) CALIFORNIA INST OF TECHNOLOGY.
XX
XX PI Kwon YT, Varshavsky A;
XX WPI; 1999-130395/11.
XX P-PSDB; AAW84351.
DR


```
QY 4657 cctgagacacctgcgttttcaactatccaggcaattgaaatctatttggagatgaaggaa 4716
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4088 cctggagcacgtgctgcttccacatccaggcaatcgaaaccttgggagatgaaggaa 4147
QY 4717 aacctctgttggagcacttcaaaataggcagcataatgctctgaaagcattaatgcagt 4776
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4148 aacctctatttggagcacttcaaaataggcagcataatgctctgaaagcgtctaatgcagt 4207
QY 4777 ttgcagttgcacagagattacactgtcctcaggtcctctgatacagaaacatctgttcgtc 4836
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4208 ttgcagttgcacagaggttacctgcctcaggtcctctgatacacaacacatctggtcgcg 4267
QY 4837 ttctacagttgttcttccataaataaaatcagaagatacaccaatgccttctgtctatag 4896
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4268 tctgtcagttatttctcacaacctgcatacagaataacacaggcctctctgtctgtg 4327
QY 4897 atctgttctatgtttgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 4956
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4328 atcttctccatgttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 4387
QY 4957 ctgtgtatctgcagcctcttctcagtttagtttcttctcataaccaccttcttcttccatt 5016
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4388 ccgtggatctgcagcgtgcacctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 4447
QY 5017 tgatcaccatggcacacatgcttccagatactacttaccagttagaca-----cag 5064
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4448 tgatcaccatggcgacatgcttccagatactcttcaacacagatacagatctgtctccag 4507
QY 5065 gcttacccttctgcaggttcaagaagacagtaggaaggtcattccgcactctcttctct 5124
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4508 ggcgcgcgttctgtgaggttgaaagagatagtgaggaggtcgtctgtcactgtcttct 4567
QY 5125 ttgcagaaatttctcaatatacaagtgggtccatttgggtgtgatatcttccgtgggtatt 5184
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4568 ttgtggaagtgtgcagcacacagacggtcctcactggtgcgtgcccggtcgttacc 4627
QY 5185 ttgtggtctcactgaaagatggataccaccttattctcgtgtgtgtgtgtgtgtgtgt 5244
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4628 ttgtggtctccttgaggaaagcagatcacaccttacctccgctgtgtgcactgttctcc 4687
QY 5245 actatttacttgggttaactcccgctgaggaaactgcataccaattctgcagaagagagt 5304
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4688 actatttacttgggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 4747
QY 5305 acagtgcactctgtagctatctatttaccatacaaaatttgttctcgtctcttccaggaa 5364
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4748 tcagtgcactctgtagctatctatttaccatacaaaatttgttctcgtcttccaggaa 4807
QY 5365 atttgggatactgtaagccttctcctcagaggttggtgtgcagatcctcgtctactaaact 5424
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4808 atttgggataccataaagccttactacagaggttggtgtgagatcctgccttactcaagt 4867
QY 5425 gtttgaagcaaaaaaacacccgtgtgcaggtaccctagaaaaaataagttttagatagac 5484
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4868 ctttgaagcgaaaaagtgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 4927
QY 5485 ttctgtatgactatagctgtcctcctgaatcaagcttctcatttccagtgccacaggtctg 5544
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4928 ttctgaggactacagctgtcttcttaaatcaggcttctcactttagtgttccacaggtctg 4987
QY 5545 cagatgtagcgaagaaacatcttctcgtccttcttcttcttcttcttcttcttcttctc 5604
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4988 cagatgtagcgaagaaacatcttctcgtccttcttcttcttcttcttcttcttcttctc 5047
QY 5605 agaactttgtccaggaaatgtgaacggggaagaggttgagcttgcattttttcaag 5664
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5048 agaactctgttccagaagaaatgtgaatggggaagaggttgagcgtgcgtttttctatg 5107
QY 5665 cacttactgtgagccgagctgtcatttctcctaaataacagagaaatgcagaggtgtcc 5724
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5108 cgcttctatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 5167
QY 5725 tgattgaagtaagcagaggtgtgtcctatccagctccttacttggatgaatatggag 5784
```

```
Db 5168 ttgttgaagggaaaaagccagaggtgtgctaccagcccttacttggatgaatatggag 5227
QY 5785 aaacagacccttgcctggaagagggaaccccttacttattctgtgagcgtgtatcgga 5844
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5228 aaacagaccagggcgttaaagagaggaacccacttacttactcgttgcgttgcgga 5287
QY 5845 agctccatttggcttgcgaacacactgcattatagaagagattgtcaggagcccaaga 5904
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5288 agctgcatttggcttgcgaacagcactgcattatagaagagattgtcgtgcggaagaga 5347
QY 5905 ctaatcaagatttatttgaattcaactggcagttactgtgagctccaactctgcctcaag 5964
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5348 ctaatcagatgctatttgaattcaactggcagttactgtgagctcagttgtgcctcaag 5407
QY 5965 acaatcaaaatgacgacagtagtaaaaggtgattcaaaaattatggaacacttctctgag 6024
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5408 acaatcatgagtgaacatcaataaaagactgatactaaaattctagagaacttcttgag 5467
QY 6025 gctgggaagatttggaggtgttcttgcctcaatgtccaggttcaactacatcaataaaat 6084
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5468 acgggggaaglatttggaggtgttcttgcctatgtccagattccacacatcaataaaat 5527
QY 6085 atttcttaagttagattgttctcaattagcaaacatgtcttcacaggaaaaa-aggac 6143
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5528 attccttaaggaaattgtcttcaattatcaaacataagcttcaagggaacaaagac 5587
QY 6144 atagatcaactctgt-----ttt 6160
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5588 atagattaattgtttatttcttagaacaactaaagaatgctgttccatccaagtgctat 5647
QY 6161 atgtgctagtatttccaggaatttattcccttccataattgtgtctatcttcttcttatt 6220
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5648 ttctgctaatttccagaaaactcttcccttcataactgctactgtcttcttcttctat 5707
QY 6221 catcacttggtagatgaagtcacgtcaaaaacagttgtgacacatttattgtgtgtttaa 6280
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5708 caccacgtgttaagaggtcacataagcaatttggacatttctccatctcgtgctaac 5767
QY 6281 tcttctcaatttgtatttgggtgttt 6307
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5768 atctctgcacotttgaatttgggtgttt 5794
```

RESULT 2
AAC86933
ID AAC86933 standard; cDNA; 6395 BP.
XX
AC AAC86933;
XX
DT 02-APR-2001 (first entry)
XX
DE Nucleotide sequence of Mouse Ubr1 protein.
XX
KW Ubr1; E3-type protein; ubiquitin system; ubiquitin-protein ligase;
KW N-end rule pathway; intracellular pathogen; Lysteria monocytogenes;
KW Yersinia enterocolitica; muscle wasting; infection; ss.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT CDS 115..5388
FT /*tag= a
FT /product= "Ubr1"
XX
PN US6159732-A.
XX
PD 12-DEC-2000.
XX
PF 11-JAN-1999; 99US-0228317.
XX
PR 02-DEC-1997; 97US-0982956.
XX

PA	(CALY) CALIFORNIA INST OF TECHNOLOGY.	
XX		
PI	Kwon YT, Varshavsky A;	
XX		
XX	WPI: 2001-090278/10.	
DR	P-PSDB; AAB31162.	
XX		
PT	Inhibiting the N-end rule pathway in mammalian cells for treating	
PT	infections and various diseases associated with muscle tissue wasting,	
PT	by inhibiting the expression of Ub1 gene	
XX		
PS	Claim 3; Column 15-28; 18pp; English.	
XX		
CC	The present sequence encodes a murine Ub1 enzyme. Ub1 is an E3-type	
CC	protein of the ubiquitin system. Specifically, it is a ubiquitin-protein	
CC	ligase. The enzyme is specific for destabilising residues exposed at	
CC	the N-terminus of protein substrates. Inhibition of the expression of	
CC	Ub1 gene in a cell results in inhibition of the N-end rule pathway.	
CC	The method is used for treatment of mammalian cells infected with an	
CC	intracellular pathogen, e.g. Lysteria monocytogenes or Versinia	
CC	enterocolitica. Inhibition of N-end rule pathway is also useful for	
CC	treating various diseases associated with wasting of muscle tissue and	
CC	infections.	
XX		
SQ	Sequence 6395 BP; 1802 A; 1376 C; 1519 G; 1698 T; 0 other;	
Query Match 70.3%; Score 4435.4; DB 22; Length 6395;		
Best Local Similarity 86.3%; Pred. No. 0;		
Matches 499; Conservative 0; Mismatches 721; Indels 69; Gaps 6;		
Qy	590 ttcaggggccgtctgaaaaagtgcgtccctgtctc-tccgaccggccacaggtttccgct 648	
Db	8 ttcaggggccgtctgaaaaagtgcgtccctgtcgcgtgcggccggccacaggtttccgct 67	
Qy	649 tgcctctggccgggggtcgccgaactcagcggcgtcagtttccctcaagatggcgacagg 708	
Db	68 agctggcgccgggggtcgccgaactcggcgcttcgtttcccttaagatggcgacgaag 127	
Qy	709 agcgtgaggtactgagagatgaaatcagcggcggaattcccccagccctcagcgtc 768	
Db	128 aga tggacggcgccgagagatggacgtcagccggagccctccctggcccgacggcg 187	
Qy	769 tggcatcttggtggatcagcaagtgaattttatctgcttcttctgcatctattggcac 828	
Db	188 cggcatcgtgggtggtcagcaagtgaattttctatctactgtttcttacctatttggcac 247	
Qy	829 aattggtgccagaaatttactttgtgtaattggaccagacttggaaagcagagggaaa 888	
Db	248 aattagtgccagaaatttattttgtgagatggaccagatttggaaagcaagaagaga 307	
Qy	889 gtgtacaaatgtcaattaccctccactcgaatggatggtaatttggagaaagatccagata 948	
Db	308 gtgtacagatgtcaattaccctccttggagtggtactatttggagaggtccggata 367	
Qy	949 ttgtcttagaaaattgaagcacagtgagcatcttcagctttgtggaggttttcaaaa 1008	
Db	368 ttgtcttagaaaattaaacacagtggtggcgttccagctgtg-gggaaggttttcaaaa 427	
Qy	1009 gtggagacaacattcttgcagggaattgtgcaattgatcccaactgtgtactctgta 1068	
Db	428 gtggagaacaacataattcctgtagggaattgtgcaattgatcccaactgtgtctgta 487	
Qy	1069 tggactgttccaggacagtggttctataaaaatcatcgttacaaagtgcatacttctactg 1128	
Db	488 tggactgttcccaagtagtgttcataaaaaccatcgtttacaagtgcatacttctactg 547	
Qy	1129 gagaggtttctgtactgtggagacacagagggcatggaaaactggccctttttgtgtaa 1188	
Db	548 gaggggttctctgtactgtggagacacagagcgtggaaaactggccctttttgtgtgg 607	
Qy	1189 atcatgaacctggaagagcagggtactataaaagagaattccagctgtccgtgtgaatgaag 1248	
Db	1688 acatggaagtggaccctgactgggaggtccatcgctctacatcgatgcaactaaagaata 1747	
Db	608 atcacgagcctggaagagcaggttactacaaaaagagagcttacattgcccattgaatgaag 667	
Qy	1249 aggttaattgtccaagccaggaataatttctctcagtgataaaaaatattgtcgtagaataa 1308	
Db	668 aggtgattgtctaagccaggaataatttccctcgttgatataaaatacattgtagaataa 727	
Qy	1309 ctatatgggaagaggaaaaaagaaactcctcctgaactccagataaaggaggaataatgaaa 1368	
Db	728 ctatatgggaagaaagaaaggaaattgctcctcctgaactgcagataaaggagaaatgaac 787	
Qy	1369 gatactattgtctcttttcaatgatgaacacacattcatatgacacgctcatatacagcc 1428	
Db	788 gatactattgtctcttttcaacgatgagcaccattcgtatgatcattgattcactacagtc 847	
Qy	1429 taaaaagactcttgaactgagctgcagagacccagtcacacactgcataccactccattgaca 1488	
Db	848 tgcagagactcttagattgcgagcttgagagggcagcagctgcacacactgcccactcgaca 907	
Qy	1489 aagaggtcgtcgggctgttaaagcgggagcttatgctgcttgcagaaagcaagaaaggaa 1548	
Db	908 aagaggtcggcggtgtcaagcaggtgtgtatgccacttgcagaaagcaagaaaggagg 967	
Qy	1549 atataaagagtcattcagaaaaatgtctctcaacatccacttcattgtagaagtattacact 1608	
Db	968 atataaagagtcactcagagaacgtctctcagcaccctccatgtggaagtgtcgcact 1027	
Qy	1609 cagagattatggtctcatcagaaaatttgcgttgcgttcttgcgttctcgtgagtaacaaaatta 1668	
Db	1028 ccgtggtattggtctaccagaaattcgtctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 1087	
Qy	1669 tgagctattcaagtacttaggcagatcttttgcgaagcagtcgcttagagaagaacctg 1728	
Db	1088 tgagctattcaagtacttagacagatatatttgcgaagcgtcgtcgttagagaagaacctg 1147	
Qy	1729 actcggagaatccctgtctcatcagaagggttaattgctttggatgcgaagctttataaag 1788	
Db	1148 gctctgaaatccctgtcttatacagcagactaatgctttggatgcgaagctttataaag 1207	
Qy	1789 gtgcccgtaagatccctcatgaattgattctcagcaggtttttttatggagatggaataca 1848	
Db	1208 gtgcccgtaagatccctcatgaattgattctcagcaggtttttttatggagatggaataca 1267	
Qy	1849 aaaaactcttgcattggaattgtgaagtattataaaacaaactgcagaaagaataataca 1908	
Db	1268 aaaaactcttgcattggaattgtgaagtattataaaacaaactgcagaaagaataataca 1327	
Qy	1909 gtgatgatcatgacagaagtatctctataactgcacttccagttcagatgtttactgttc 1968	
Db	1328 gcgacgaccacgagagaagcatctccataaccgcctgtccgtgcagatgctcacccgtcc 1387	
Qy	1969 ctactctggtcgcagactcttattgaagagcagaaatgttatctctgtcattactgaaactc 2028	
Db	1388 cgaccttggcccgccactcttattgaagagcagaaatgttatctctgtcattactgaaacgc 1447	
Qy	2029 tctagaagtgtttacactgagctacttgacagcagaaataataactccagggttata 2088	
Db	1448 tctagaagtgtttacactgaaacttggacagggaacataataactccagggttata 1507	
Qy	2089 gccagacaaaattgggaagagtatgacagtaatatatgacactaaagtatatcctgatca 2148	
Db	1508 gccagacaaaattgggaagagtatgacagtaatatatgacactaaagtatatcctgatca 1567	
Qy	2149 gaaaaccccaataatggacagaaagattaagaatgcagttccttgcaggttttcgactctt 2208	
Db	1568 gcaagcgtctcatatggacagaacgattaagagcagcttctggaaggtttcccggtctt 1627	
Qy	2209 ttttaagaattcttactgtatgcagggaatgaagaaatccgagacagctgtgggcaac 2268	
Db	1628 tctgaagattcttactgtatgcagggaatgaagaaatccgagacagctgtgggcaac 1687	
Qy	2269 acatgaaatggatcctgattgggaggtcctcattgtctatcacagatgcaattgagaata 2328	
Db	1688 acatgaaatgggaccctgactgggaggtcctcattgtctatcacagatgcaactaaagaata 1747	

Qy 2329 ttttaactcatgtttccaagagtggtgtgtcttgtgtgatgaagaactcttacttctgtggtctata 2388
Dbb tttgtctcatgtttccaagagtggtgtgtcttgtgtgatgaagatcttacttctgtggtctata 1807
Qy 2389 aagaatgtccaaaagctgtgtgatgaggtgtcaggtaccaggtttcatactatagtagcaagacag 2448
Dbb 1808 aagaatgtccaaaagctgttaatgagggtcaggtacaataatttcatagttccagttaccaagacag 1867
Qy 2449 tagtacaatcgtgtggacatagttgtgaaacaaaagtcctacagagtatctgagagatcttg 2508
Dbb 1868 tagtgaatgtgcgtcatagttcgtgaaacaaaatcctacaaggtgtcaggaaccttg 1927
Qy 2509 taagcátacatctgcacactctctagggaccttgcctgtcttcatgtacgttttaagcaggc 2568
Dbb 1928 taagcátacacactctcagacacttgcctgtcttcatgtacgttttaagcagac 1987
Qy 2569 tgggtcgtgttcaagactgcataaattgtgtcttttggaggaccttcaagtagagatc 2628
Dbb 1988 taggtgcta ttcaagactgcataaattgtgtcttttggacagcttcaagtagaggtcc 2047
Qy 2629 tagtggaaatcctttacgttctgcgtgtgtgtgtgcccagggtgtgtcgtgagatgtggc 2688
Dbb 2048 tgggtgagtaaccgctgcgtgcgtgtgtgtgtgtgtcaggtgtgtgtcgtgagatgtggc 2107
Qy 2689 gaagaaaatggactgtctcttatttagccaggtgtgttttattaccagatgtttaagtgcagag 2748
Dbb 2108 gaagaaaagggctctcactcatcagccaggtgtttctattatcaagatgtttaaatgcaagg 2167
Qy 2749 aagaaatgtatgataaagatatcatcgtctcagatgtgtggtgcattttaaagtgcacca 2808
Dbb 2168 agaaaatgtacgataaagatatcatcgtctcagatgtgtgagcactataatggtgcacca 2227
Qy 2809 ataagttctgttactgttacttcagaggtatgaacttgcgaggtctttaacaagacca 2868
Dbb 2228 acaagttctgttactgttacttcagagatatgaacttactgtatgttttaacaagacca 2287
Qy 2869 tatctcaaaagaccaggtattgattaaacaataatacactaataagaagaatgtctc 2928
Dbb 2288 tatccaaaaagaccaggtattgattaaacagataatacactaataagaagaatgtctc 2347
Qy 2929 aggtccctcatctatattgtgtgggtgagcgttatgtacatgaggtgggaaatgtgaccaaaag 2988
Dbb 2348 aggtccctcatctatattgtgtgggagacgttatgtacatgaggtgggaaatgttaccagag 2407
Qy 2989 aagaggtccaatgagagaaatcattcacttgccttgcattgaacctcagccatgccacacagtg 3048
Dbb 2408 aggaggttaaatgagagagattactcacttacttgcatttgagcccatgccacacagtg 2467
Qy 3049 ccattggccaaaaatttacctgagaatgaaataatgaaactggcttagagaatgtcataa 3108
Dbb 2468 ccattggccaaaaactacctgagaacgaaaaataatgaaactggcttagagaatgtcataa 2527
Qy 3109 acaaaatggccaatatttaagaacacaggtgtatcaagccatggaagttatgaacataaag 3168
Dbb 2528 acaaaatggccaatatttaagaacacaggtgtgtcggccatggagtttatgaattgaaag 2587
Qy 3169 atgaatcactgaagacttcaatattgtacttttacttactcactcacaacccagcatagca 3228
Dbb 2588 atgaatcactgaagacttcaatattgtacttttaccattatttctaaacacacagcatagca 2647
Qy 3229 aggtgaacataatgcagaagaaagggaaacaaagaaacaaagataagagcatgtccgc 3288
Dbb 2648 aggtgaacataatgcagaagaaagggaaacaaagaaacaaagataagagcatgtccgc 2707
Qy 3289 caccacacactctgaattctgcctgtcttccagcaaggtattaaaccttctcaactgtg 3348
Dbb 2708 cgcacactctccagagttctgcctgtcttccagcaagtagtcaacctgtcagctgtg 2767
Qy 3349 atatcatgtatgatacattctcagaccgtattttagcgggcaatagacacagattctcaact 3408
Dbb 2768 atgttatgatacattctcagaccgtattttagcgggcaatagacacagattctcaact 2827

Qy 3409 tgtgaccgaagaggatggtcctcaaaatggtcttctcatattcttggcatgggttttactagaag 3468
Dbb 2828 tgtgaccgaagaggatggtcagatggcgttccatattattggcactgggtttctggaag 2887
Qy 3469 aagaacacacagcttcaaaaagctcctgaagaagaagtaacatttgacttttatacataag 3528
Dbb 2888 aagaacacacagcttcaaaaagctcctgaagaagaagtggtcttggacttttaccataaag 2947
Qy 3529 ctccaagatttgggaagtctagccat-----gaatatacaaaatgcttttggaaaaac 3579
Dbb 2948 ctccaagatttgggaagtctagccatgaatgctcagaatatatacaaatgctcttggaaaagac 3007
Qy 3580 tcaaaaggaattcccccagtttagaagccagaagacatgataaocgtggatctcagatgt 3639
Dbb 3008 tcaaaaggaattcccccagtttagaagccagaagacatgataaocgtggatctcagatgt 3067
Qy 3640 ttgacacagtgaaagcgtattaaagaaaaatcttctttaaattgtagcaaacacacatcagat 3699
Dbb 3068 ttgacacagtgaaagcgtattaaagaaaaatcttctttaaattgtagcaaacacacatcagat 3127
Qy 3700 cggaaatttataagaatgatgagattactcatgataaaagaaaaagcagaacgaaaaagaa 3759
Dbb 3128 tggagtgcatlaaagagtgaaggattactcatgataaagaaaaagcagaacgaaaaagaa 3187
Qy 3760 aagctgaagcgtcgtcaggtacatcgcagaagacatcgtcagatgctcgtccttaacaga 3819
Dbb 3188 aagctgaagcgtcgtcaggttccatcgcagaagacatcgtcagatgctcgtccttaacaga 3247
Qy 3820 aaaaactcatgaaaactcataaaactcatgataaactacatcagaaaatgctcgtggaaaag 3879
Dbb 3248 aaaaactcatgaaaactcataaaactcatgataaactacatcagaaaatgctcgtggaaaag 3307
Qy 3880 aagattccattatggaggaagagagaccccccagcagtcagtgactactctagaaattgctt 3939
Dbb 3308 aagattccattatggaggaagagagaccccccagcagtcagtgactactctagaaattgctt 3367
Qy 3940 tgggtcctaaagcgggtccatctgttacttgaagaagaggtgctgacgtgcactccttgcc 3999
Dbb 3368 tgggccttaaacggggccggctgttacgcgaagaggtgctgacgtgcactccttgcc 3427
Qy 4000 aagaagacacagaggtgaaaaatagaaaaatagccatgattattatcgccgtcgtctgcaga 4059
Dbb 3428 aagaagacacagaggtgaaaaatagaaaaatagccatgattattatcgccgtcgtctgcaga 3487
Qy 4060 aatctacttgccttaacccagcagcaggggaaaaaccccatagaactctcagaggaagccctag 4119
Dbb 3488 aatccacgccttaacccagcagcagggaggaagcctgtggaaccttaggggaacacactgg 3547
Qy 4120 acccaacttttcattgatccagacttggcataatggaaacttataoaggaagctgtggtcatg 4179
Dbb 3548 accctcttttcattgatccagacttggcataatggaaacttataoaggaagctgtggtcatg 3607
Qy 4180 taatgcacagcagtgctgtggcagaagatttttgaagcttatacagctgagctctcagcagc 4239
Dbb 3608 taatgcacagcagtgctgtggcagaagatttttgaagcttatacagctgagctctcagcagc 3667
Qy 4240 gcatttcatgttgaccttttgcacttggaaaatggagaatatcttttgcctcttgcacaaat 4299
Dbb 3668 gcatttcatgtagacctgtttgaccttggagagcggcgagcacctatgtcccgtctgcaagt 3727
Qy 4300 ctctgtgcaatactgtatccccattattctcttggcaacttcaaaagataaaacagtgaga 4359
Dbb 3728 ctctctgcaacactgtatccccattctcttggcaacttcaaaagataaaacagtgaga 3787
Qy 4360 atgcagatgctctgtctcaacttttgaoccttggcaggttggatacagactgttcttgcca 4419
Dbb 3788 atgcagaggtctctgctcaacttttgaoccttggcaggttggatacagactgttcttgcca 3847
Qy 4420 gaatatcaggttataataataagacatgctaaagggaaaaaac---caattcctatttct 4476
Dbb 3848 gaatatcaggttataataataagacatgctaaagggaaaaaac---caattcctatttct 3907
Qy 4477 ttaatacagggaatggagatctctacttttggagttccattccactcctcgtgatttttggcgttg 4536

Db 3908 taaacaaaggatgaggatcaactcttgagtttcaatccatcctgagtttgagttc 3967
Qy 4537 agtcttcgattaaattacaaatagatcaaggaaatggttattctcttgccacaacaa 4596
Db 3968 agtcttcgtagaataattcaaatagatcaaggaaatggttattctcttgccacaacaa 4027
Qy 4597 ttatagaaattggattgaaatgcaactgatgaaaggatccctcgagtcctccatgctga 4656
Db 4028 ttacagaattggcctgaaagtgcctcctgatgaactagaccacagagtgcccatgatga 4087
Qy 4657 cctgagcaactgcgcttctactatccaggcaattgaaattctatctggagatgaagaa 4716
Db 4088 cctggagcacgtgctgtccactccaggcaattcgaaacactgttggagatgaaggaa 4147
Qy 4717 aacctctgttggagcaacttcaaaataggcagcataatggtctgaaagcattaatgcaat 4776
Db 4148 aacctctatttggagcaacttcaaaatagacagcatagcggtctgaaaggcctaatgcaat 4207
Qy 4777 ttgcagttgcacagagattacacctgtcctcaggtcctcgatatacagaacacatctggtctgc 4836
Db 4208 ttgcagttgcacagaggtaccctgcctcaggtcccttgatatacacaacacatctggtctgcg 4267
Qy 4837 ttctatcagttgttcttctactaaataaaatcagaagataacacatgccttctgtctatag 4896
Db 4268 tctgtcagttattcttctactaacctgcaatcagaaaaatcacaccagccttctgtctgtgg 4327
Qy 4897 atctgttctatgttgggtggtcgtgtgttagcatctccatcctctgtattggagtgacc 4956
Db 4328 atcttctccatgttctgtcgccagctcttagcgttcccatcctctgtattggagtgaca 4387
Qy 4957 ctgttgatctgcagccttcttcagttagttcttctctataaacacaccttatctctccatt 5016
Db 4398 ccggtgattgcagcgtcgccactagttcttctatataaacacaccttatctctccatc 4447
Qy 5017 tgaataccatggcacacatgcttcagatactactactacagtagtaca -----cag 5064
Db 4448 tgaataccatggcacatgcttcagatactactactataacacagatacagatctgtctccag 4507
Qy 5065 gccatccctgtcaggttcaagaagacagctgaaggctcattccgcattcttctct 5124
Db 4508 gggccgcctgtcgtgaggggtgaagagatagtgaggggctcgtgtgcatactgcttct 4567
Qy 5125 ttgcagaataattctcaatatacaagtggtcctcattgggtggtgatctcctggctggtatt 5184
Db 4568 ttgtggaagtgcgcagcacacagacggtcctcactgggtgcggtcctccggctggtacc 4627
Qy 5185 tgtgggtctcactgaagaatggcatccccccttatcttgcgtgtgctgattgttttccc 5244
Db 4628 tgtggctcctcagggaacggcatcacccttacctcctcgtgtgctgcaactgctttcc 4687
Qy 5245 actatttacttgggttaactccgcctgaggaacgcataccacaaattctgcagaaggagat 5304
Db 4688 actatttacttgggtagctcgcctgaggaactgttggcaattctgctgaaggagaat 4747
Qy 5305 acagtgcactctgtagctatctatttacctacaaaattgttctcgtcttccaggaat 5364
Db 4748 tcaagtgcactctgtagctatctatttacctacaaaattgttctcgtcttccaggaat 4807
Qy 5365 attgggatactgtaaggcccttgcctcagaggtggtgtgcagatccctgccttactaaact 5424
Db 4808 attgggataccataaaggcccttactacagaggtggtgtgcagatccctgccttactcaagt 4867
Qy 5425 gtttgaagcaaaaaaacacggtgctagggtaccctagaagaaaaatagtttgatagagc 5484
Db 4868 ctttgaagcagaaagtgtgtgtcaggtaccctagaagaaaaatagtttgatagagc 4927
Qy 5485 ttctctgactatagctcctcctgaatcaagcttctcatttcaggtgcacaggctctg 5544
Db 4928 ttctctggagactacagctgcttcttaatacagggtcttctcatttaggtgtccacggctctg 4987
Qy 5545 cagatgatgagcgaagcaactcctcctctgcttttctgtggggctatactatgttctc 5604

Db 4988 cagatgatgagcgaagcatcctgtctcctgtcttcttctgtgtgggcaactcctgtgtctc 5047
Qy 5605 aqaacatttgcctgcagggaatgtgaacggggaagaggttgagcttgcatatttcaacg 5664
Db 5048 aqaacatttgcctgcagggaatgtgaatggggaagaggttgagcttgcatatttcaacg 5107
Qy 5665 cacttcaacttggagcggagctgtcattttcctaaaaatcagaagaatgcgagtggtcc 5724
Db 5108 cgtcttcatttgggtgctgaggtcgtcattttctaaaaatccgagaatcgaggtgtgctc 5167
Qy 5725 tggttgaaggtaaaagccagaggtgtgctctatccagctccttacttgatgaatatagag 5784
Db 5168 tggttgaaggtaaaagccagaggtgtgctctacccgccccttacttggatgaatatgag 5227
Qy 5785 aaacagaccctgcctgaagaggggcaacccctcatttctctgtgagcgatcagga 5844
Db 5228 aaacagaccgaagggtlaaagagaggaaacccacttcttctcgggagcggtatcagga 5287
Qy 5845 agctccatttggctggaacaaactgcattatagaagagattgctagagccaaagaga 5904
Db 5288 agctgcatttggctggaacagcactgcattatagaagagattgctcggagccagaga 5347
Qy 5905 ctaactcagatgttatttggattcactgcagtgctactgtgagctccaaactctgcctcaag 5964
Db 5348 ctaactcagatgtatttggatttaactggcagttactctgagcttcagttctgcctcaag 5407
Qy 5965 acaatcacaaatgacacagatgaaggctgattcaaaaattatgaaaaacttctcagag 6024
Db 5408 acaatcagtgagcatcaataaagactgactcaaaaattctagagaacttctcagag 5467
Qy 6025 gctgggaagatgtgaggggtcttctgcctatgctcaggttccacttaccatcaataaaat 6084
Db 5468 acgggggaagatgtgaggggtcttctgacccatgctccagattcacacataataaaat 5527
Qy 6085 attcttaatgagatgttcttcaattagcaaaacatatgtctccacaggaaaaa -agagc 6143
Db 5528 attccttaaggaatattgcttcaattatcaaacataagcttcaagggaagaaacagac 5587
Qy 6144 atagatcaactgt -----ttt 6160
Db 5588 atagattcaatgttttatgttctagaacactaaagaaatgcttcttccaaagtctctat 5647
Qy 6161 atgtgctagatttccaggaatttattcccttcaataatttgcctcatttctcattttatt 6220
Db 5648 ttctgctaatatttccagaaaactccttctccctcaactgcctgagttcatttctat 5707
Qy 6221 catccacttggtagatgaagtcacgtcaaaacagttgttagacattttatgtgtgggttaac 6280
Db 5708 caccacactggttaatgaggtcacattgaagcatttgggacatttctccatctggctaac 5767
Qy 6281 tcttctgcaatttctgatttgggttt 6307
Db 5768 atctctgcaccttctgatttgggttt 5794

RESULT 3

AAH14878

ID AAH14878 standard; cDNA; 2550 BP.

XX AAH14878;

XX 26-JUN-2001 (first entry)

XX Human cDNA sequence SEQ ID NO:12731.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX

Db 668 aagatacagaactccacactccgggtgttgctccattctgcctctgtgttgcaagcctgg 727
Qy 3331 ttaacctctcaactgtatgatcatgtatcatcttcagagaccgtatttgagcggca 3390
Db 728 ttaacatttgcagtcagatgtcatgttgcatactgggaacaattctgcaatggcgtg 787
Qy 3391 tagacacagatttaacttctgtgacccgaagggtgctcccaaatgcttttcaattctgg 3450
Db 788 tgaacataatggatagctgtgtcagagtcctcagtgcaagggtgttacatttaattg 847
Qy 3451 cattgggtttactagaagaagaacagcttcaaaaagctcctgaag---aagaagtaa 3507
Db 848 gcatggcactacaagaagaacacacatttagaagatgtcacggagagcatgtagtaa 907
Qy 3508 catttgactttatcataaggtcttcagatttggaagtctcagccatgaat-----a 3558
Db 908 catttaccttcacagaagatatacaaaacctgtgtgaagcgcacaaaatttctctagca 967
Qy 3559 tacaatatgttttggaaaaactcaaaaggaattccccggttagaaggccagaagacatga 3618
Db 968 tactagtatgtggaaacactcaaaaatgctcctacctcagtagaagtcacaaaagacatga 1027
Qy 3619 taacgtggaactacatgatttgacacagtgaaagcgaatgaagagaaaaattctgtttaa 3678
Db 1028 ttcgggtggaattgaagactttcaatgtcttcaaaaagatgaaggaga-----gttcaac 1081
Qy 3679 ttgtagcaaccacatcaggatcggaattcttaagaatgatgatatctactcatgataaag 3738
Db 1082 ctaccagtcctgtggcagacagagaacacataatggaagagatttcaagggaacaaag 1141
Qy 3739 aaaaacagacaagaaaaagcgtgaagcgtgtagctgtagctacatcgcagaagaatcatgg 3798
Db 1142 acaagctgagagaagaaaaagacagatttcgcaactgctgcagagaaaagatcatgg 1201
Qy 3799 ctcaagatgctccttacagaaaaaacttcaaaactataaactcaactcatgtatgaacata 3858
Db 1202 ctcaagatgctgaaagcagcggcattttattgtatgaaaaaagaactcttcagcaga 1261
Qy 3859 catcagaaaatgcctgggaagaagattccattatggaggagaagacacccacagatca 3918
Db 1262 cattagaactggatgcctcaactctgtgttcttgata-----tagcccttggtctt 1315
Qy 3919 gtgactactctagaattgtttgggttcttaaacgggggtccatctgttactgaaaagagg 3978
Db 1316 cagatatgaacttacagcactgggcccgcacaaaactcaggttccctgaacaaagacaa 1375
Qy 3979 tgcgtgagtgatcccttggcagaagaacagagggtgaaaaatagaataaataatgccatgg 4038
Db 1376 tegtacatgtaattgttcaagaggagcaagaagttaagtgaagtcagagcaggcaatgg 1435
Qy 4039 tattatcggtctgttccagaagaacttactgccttaaccagcagacaggggaaaaacccatag 4098
Db 1436 tcttggcagcatttgttcagagatacaactgtattatacaaaaacagaagtaaaatttttc 1495
Qy 4099 aacttcagagagaagccctagaccacttttcatggatccagacttggcatatggaactt 4158
Db 1496 aagatccagaaaaatat---gatccattattcatgcaccctgatctgtcttgtggaacac 1552
Qy 4159 atacagaagcgtgtgtcatgtaatgcacagcagtggtctggcagaagaatttttgaagctg 4218
Db 1553 acactagtagctgtggacacattatgcattgcccattgttggcaaggtattttgatccg 1612
Qy 4219 tacagctgagctctcagcgcgcatcattgttgacctttt-----tgacttgg 4266
Db 1613 ttcaagctaaagacagcgaagcaacagagataacgcttacaacagcgtatgtatgag 1672
Qy 4267 aagtgagagaatactttgccccttttggcaaatctctgtgcaatctgtgacactgtatcccat 4326
Db 1673 aaacagagaattcctttgccccttttggcaatgtgaatgcttgagtaataactgtattctctgc 1732
Qy 4327 ttctcttgcaacctcaaaagataaaacagtgagaatgcagatgctctgtcctcaactttga 4386

Db 1733 tgcttt-----ctccagaagaataatttttaaacacaggtttaaaattttttcagaccaacaa 1786
Qy 4387 ccttggcagcgttggatcacagactgtttctgcccagaatatcatcaggtttaataataagacatg 4446
Db 1787 atctgactcagtggttagaacaatactcagcaataaaagcattacaggtttcttagga 1846
Qy 4447 ctaaaaggagaaaacccaattcctattttctttaaatacaaggaaatgggagattctactttgg 4506
Db 1847 aagaagaagtagtctcctaataatgcctctacaagaattcagaaaaatgttgatgaattac 1906
Qy 4507 agtctcattccatcctcaggttttggcgttgagcttcgattaaatatcaaatagatca 4566
Db 1907 agtctcctgaagggttcaggcctgatttctgtcctaagaacctcttctctgagagcata 1966
Qy 4567 aggaataggttattctcttgcacacaatttatagaatttgattgaagtgccacctg 4626
Db 1967 aagaatgctaacgacatttggaaactgtacctacaagggtgggactaaaagttcatcca 2026
Qy 4627 atgaaggagatcctcagatccccatgtcgaactggagcactgcgtcttcaactccagg 4686
Db 2027 atgaaggagatcctcgtgtcccataatgtgttggggtagctgcgtacacacatccaaa 2086
Qy 4687 caattgaaaatctatttgggagatgaagaaaacctctgttctgagcacttcaaaaatagcc 4746
Db 2087 gcatagaagaatttggatgtagaataaaacctgttctgtcccttacccttgcagac 2146
Qy 4747 agcataatggctgaagcatttaatacagtttgcagttgcacagagatcacctgtcctc 4806
Db 2147 tggatgactgtcttgggtcagcagatattgcccagcacacactggacagtggtcatcag 2206
Qy 4807 aggtcctgatacagaaaaacatcgttgcgtctctctatacagttgttcttcttaacataaaat 4866
Db 2207 ttctagtgtagcaagacattttttaaacttttgcactcactggtgcctaaatgacagcc 2266
Qy 4867 cagaagatacacacgcctctgtctatagatcgttttcatgttttctgttgggtggtgtgtg 4926
Db 2267 atgggaacttccatgcatactagatattgacatgtttcatttatgttgggtggtgtgtg 2326
Qy 4927 tagcattcccatccttctgtatttgggatgacctgttgatcgcagccttctcagttagtt 4986
Db 2327 ttgcatttctcgtgtgcagt-----tcaggatttttcagggatcagcc 2371
Qy 4987 ctctctataaccacactttatctctcttccatttgatcacaccatggcacacactgttcaagatac 5046
Db 2372 ttggcactggagaccttcacatttccatctgtgttactatggcacatcatcacagatct 2431
Qy 5047 tacttacagtagaacac---aggcctacccttgcctcaggttcaagaagacagtgaaagg 5103
Db 2432 tacttacctcatgtacagaagaaatggcatggatcaaaaaatcccccttctggaagaag 2491
Qy 5104 ctcatccgcacatcttcttcttgcagaaatttctcaatatacaagtggtcctcattgggt 5163
Db 2492 aatcagcagttctcgtttgtataaaaaacttccacagtatcacgggaagtgccttgaaag 2551
Qy 5164 gtgatattcctgtgtgttatttgggtctcactgaagaatgggcatcaccccccttatcttc 5223
Db 2552 aaataccatccggctggcactcgtggaggagtgacagcgtggaatcactccttctctga 2611
Qy 5224 gctgtgcagctgttttctcaactatttacttgggttaactccgcctcaggaactgcata 5283
Db 2612 agtcttgcgttatttttcttacttaaatggagtcttctcccccacccagatccaag 2671
Qy 5284 ccaattctgcagaagagagtagacagtgacactctgtagctatctatcttaccctacaaatt 5343
Db 2672 t---tcttggaaacagccatttgaacatttattgtatgtatttcttccctacacaaaccc 2728
Qy 5344 tgttctcgtctcttccaggaaatatttgggatactgtaaggcccttgcctccagaggtgtgtg 5403
Db 2729 tcaatttgccttttcaagaaaaatagtgagataatgaattcactgattgaaagtgtgtgcc 2788
Qy 5404 cagatccttgccttactataactgtttgaagcaaaaaaacacccgttggcaggtacccctagaa 5463
Db 2789 gtaacagtgaaagttaaaagatactagaaaggtgaaagagatgctataagaatatccaagag 2848

QY 3391 tagacacagattctaaacttggaccgaagggtgctcccaatggcgttttccattattcgg 3450
Dbb 843 tggacaataatgatagctgctgctgagagtcacgtgcaagggtgttccatttaattg 902
QY 3451 cattgggtttactagaagaagaacacagcttcaaaaagctcctgaag---aagaagtaa 3507
Dbb 903 gcatggcactacaagaagaacacacatttagaagatgcacggaagcatgtagtaa 962
QY 3508 catttgacttttatcataagcttcaagattggaagttcagccatgaat-----a 3558
Dbb 963 catttaccttccctcagaagatcacaaaacctggtgaagcgcacaaaatttccctagca 1022
QY 3559 tacaaatgcttttggaaaaactcaaaaggaaattcccccagtttagaagccagaaagacatga 3618
Dbb 1023 tactagctatgctggaaacactacaaaagtctccctaccttagaagttcccaagacatga 1082
QY 3619 taacgtggatacttcagatgtttgacacagtgaaagcgattaaagaaaaatcttgtttaa 3678
Dbb 1083 ttcggtggatattgaagacttttaatgctgttaaaaagatgagggaga-----gttcac 1136
QY 3679 ttgtagcaaccacatcaggatcggaatcttattaagaatgatgagattactcatgataaag 3738
Dbb 1137 ctaccagttccgtggcagagacagaaggaaacacataatggaagagagttcaagggacaaa 1196
QY 3739 aaaaagcagaacgaaaaaagaaagctgaaagctgtaggctacatccagacagaaatcatgg 3798
Dbb 1197 acaaaactgagagaagaagaagacagagattgccagactgcgcagagaagaaatcatgg 1256
QY 3799 ctoagatgtctgcttcacagaaaaaacttcattgaaactcataaactcatgtatgacaata 3858
Dbb 1257 ctccagatgctgaaatgcagcgccattttattgtatgataaaacaaagaactcttccagcaga 1316
QY 3859 catcaaaaatgctgggaagaagaattccattatgaggaagaagacaccccagcagtcga 3918
Dbb 1317 cattagaactgagtcgctcaacctgctgttcttctgataca-----tagccctgtgctt 1370
QY 3919 gtgactactctagaattgtttggctccaaacggggtccatctgttacttgtaaaaggagg 3978
Dbb 1371 cagatatgacacttacagacactgggcccgcacaaactcaggttcctgtaacaaagacaa 1430
QY 3979 tgctgagctgcatcctttgccagaagaacagaggtgaaaaatagaaaaataatgccatgg 4038
Dbb 1431 tcgttacatgtatattgttcaagaggagcagaagttaaagtggaaagcagggcaatgg 1490
QY 4039 tattatcgccgtgtgtccagaataactactgccttaaccagacacaggggaacacccatag 4098
Dbb 1491 tcttggcagacttgttcagagatacaactgtattatcaaaaaacagaagttaaaattattc 1550
QY 4099 aactctcaggagaagccctagaccacttttcatggatccagacttggcatatggaaactt 4158
Dbb 1551 aagatccagaataat---gatccattattcatgcaccctcattctgtcttgggaacac 1607
QY 4159 atacaggaagctgtggtgtcatgttaatgcacgcaggtgtgctggcagaagttatttgaagctg 4218
Dbb 1608 acactagtagctgtgggcacattatgcatgcccattgttggcaaaagttatttatttcg 1667
QY 4219 tacagctgagctctcagcagcgcattcatgttgaactttt-----tgacttgg 4266
Dbb 1668 ttcaagctaaagacagcgaaggcacaagatattacgcttcatatacagactatgatgtag 1727
QY 4267 aaagtggagaatacttttggccctcttggcaaatctctgtgcaatactgttatccccattta 4326
Dbb 1728 aaaaaggagaaattccttggccctcttggaaatggtttagtaactgttattctctcgc 1787
QY 4327 ttcctttgcaacctcaaaagaataacagtgagaatgcagatccttctgcacattttga 4386
Dbb 1788 tgcctt-----ctccaagaataatttttaacaaacaggttaaatttttcagaccaccaa 1841
QY 4387 ccttggcagctgtggatcacagactgttcttggcagaataatcatcaggtttataataaagacatg 4446
Dbb 1842 atctgactcagtggaattagaacaaatctcagcaaaataaaagcattacagtttcttagga 1901
QY 4447 ctaaaaggaaaaaccccaatttcttattttcttaatacagggaatgagagatttacttctgg 4506

Dbb 1902 aagaagaagtagtctctataataatgcctctacaagaattcagaaaaatgttgatgaattac 1961
QY 4507 agttccattccatcctcagagtttttggcgttggagttcttcgatttaataatctcaaaatagacatca 4566
Dbb 1962 agtccctgaagggttccagccctgatttgcctcctaagatcccttattcttgagagcataa 2021
QY 4567 aggaataatgttattctcttggccacaaatattatagaatttgattgaaagtgccacctg 4626
Dbb 2022 aagaataatgtaacgacatttggaaactgtaacctacaaggtgggactaaaaggttcatccca 2081
QY 4627 atgaagaagatcctcagagtcctccatgctgaacctggagacactgcgcttttcaactaccag 4686
Dbb 2082 atgaagaagatcctcgtgttcccataaattgttggggtagctgcgtacacattccaaa 2141
QY 4687 caattgaataatctattgggagatgaaggaaaacctctgtttggagcacttcaaaaataggc 4746
Dbb 2142 gcatagaagaataatttgagtgatgaagataaaacctattgttggctcttacccttgagac 2201
QY 4747 agcataatggtctgaaagcatttaatgagtttgcagtttgcacagagagattaccctgctc 4806
Dbb 2202 tggatgactgtcttaggttcattgacgagatttggcgcagacactggacagtggcacag 2261
QY 4807 aggtcctgatcacagaacatcctggttgccttctatcagttgttcttcttaacataaaat 4866
Dbb 2262 ttcagtggtgcaaggacattttttaaacttttgcataacttgcactggtgcctaatgacagcc 2321
QY 4867 cagaagacacacacatgcctctctgtctatagatctgtttcatgttttgggtgggtgctgtgt 4926
Dbb 2322 atgaggaacttccatgcatatttagatattgacatgttctattatttgggtgggtgctgtgc 2381
QY 4927 tagcattcccatcctgtgtattgggagtgacctgttgcagcttcttgcagcttcttgcagtttagtt 4986
Dbb 2382 ttgcatcttccgtgtgagtg-----tcaggatttttcaggagatcagcc 2426
QY 4987 ctctctataacacacacttattcttccatttgcacacatggcacaactgcttccagatc 5046
Dbb 2427 ttgycactggagaccttccatttccatctgttactatggcacaactacacagatct 2486
QY 5047 tactacagtagacac---aggcctacccttgcctcaggttccaagaagacagtgaaagag 5103
Dbb 2487 tacttactcctatgacagaagaagaatggcatggatacaagaatacccccttggaaagaag 2546
QY 5104 ctaattccgcactcttcttcttggcagaataattctcaatatatacagaatggctccattgggt 5163
Dbb 2547 aatcagcagttcttgccttggataaaaacacttccacagtatatacgggaagtgccttgaag 2606
QY 5164 gtgatatctcctgctgtatttgggtgtcctcactgaagaatggcacaaccccttattcttc 5223
Dbb 2607 aatcaccatccggtggcatctgtggagagtgctcagagctggaatcatgcttcttctga 2666
QY 5224 gctgtgctgactgttttccactatttacttgggttaactccgctgaggaactgcata 5283
Dbb 2667 agtgttctgtttatttttcttacttacttaaatggagtcttccccccacacacacag 2726
QY 5284 ccaattcgcagaaggagagtacagtgcaactctgtagctatctattcttatacacaatt 5343
Dbb 2727 t---tcttggaaacagccatttgaacatttatgtatcttcttccctaccacaacac 2783
QY 5344 tgttctgcttctccaggaataattgggatactgttaaggcccttgcctcagaggtgtgtg 5403
Dbb 2784 tcatgttgccttttcaagaataatgagataatgaatttcactgattgaaagtgtgtgc 2843
QY 5404 cagatcctgcttactaaaactgtttgaagcaaaaaaacacccgtggtgaggtaccctagaa 5463
Dbb 2844 gtaacgtggaagttaaagataatcagaaggtgaaagagatgctataagatatccaaag 2903
QY 5464 aagaataatgtttagagagcttctctgtagtactatagctgcctcctcagaaatcaagttctc 5523
Dbb 2904 aatctaacaataataaaaccttccagaggtattacagcagcctcatttaatacaagcattca 2963
QY 5524 atttcaggtgccacggtgtgcagatgatgagcgaagacatcctgtcctcgtcctttct 5583

Dbb 2964 atttctcgtcgcgcgaataacaggtggtgataagacagagccccaactctgtgccttgtgt 3023
Qy 5584 gtggggtctactatgttctcagaaacatttctccaggaataattgtgaacggggaagagg 5643
Dbb 3024 gcgaatctcgtctgtccagatttactgtctccagactgaactggaaggagagtg 3083
Qy 5644 ttgagcttgatttttccagcacttcaactgtgagccggagctgtcatttctctaaaaa 5703
Dbb 3084 taggagctgcagcgtcacacttactctgtgctcgtgagtggtgagcttctcctgagag 3143
Qy 5704 tcagagaatccgagtggtctcgttgaagttaagataaagccagagaggtgtccctatccagtc 5763
Dbb 3144 tcaaggatgcaggtgtattttagctggcaaaccaaggctgttttattctctc 3203
Qy 5764 ctacttgatgaatatggagaacagaccctggcctgaagaggggcgaaccccttcatt 5823
Dbb 3204 ctactctgatgactatgggagaccagcagagactcagagcggggaatacttctacatt 3263
Qy 5824 tatctcgtgagcgtatcggagctccatttggctcgtggcaacacactgcattatagaag 5883
Dbb 3264 tatgaaagcgtattcaagaagattcagaagctcgtggcaccacacacagtgctcacagagg 3323
Qy 5884 agattgtagagcgaagagactaatcagatgttatttggattcaactggca 5935
Dbb 3324 aaattgacatgcacaggaagccaatcagacactggttggtgactggca 3375
RESULT 8
AAA02411
ID AAA02411 standard; cDNA: 733 BP.
XX
AC AAA02411;
XX
DT 19-MAY-2000 (first entry)
XX
DE Human colon cancer cell line polynucleotide sequence SEQ ID NO:2402.
XX
KW Human; colon cancer; tumour; diagnosis; gene expression product;
KW probe; detection; cancerous state; metastasis; identification;
KW breast cancer; oestrogen receptor-positive breast cancer; therapy;
KW oestrogen receptor-negative breast cancer; lung cancer; ss.
XX
OS Homo sapiens.
XX
PN WO9958675-A2.
XX
PD 18-NOV-1999.
XX
PF 13-MAY-1999; 99WO-US10602.
XX
PR 14-MAY-1998; 98US-0085426.
PR 15-MAY-1998; 98US-0085537.
PR 21-MAY-1998; 98US-0085696.
PR 21-OCT-1998; 98US-0105234.
PR 27-OCT-1998; 98US-0105877.
XX
PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEO INC.
XX
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
PI Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
XX
DR WPT; 2000-126369/11.
XX
XX Polynucleotide library used to determine cancerous states of mammalian
PT cells -
XX
PS Claim 1; Page 958; 1097pp; English.
XX
XX AAA00010 to AAA02716 represent polynucleotides isolated from cDNA
CC libraries constructed from human colon cancer cell lines. The present

CC invention also describes a method of detecting differentially expressed
CC genes correlated with a cancerous state of a mammalian cell, comprising
CC detecting at least one differentially expressed gene product in a test
CC sample derived from a cell suspected of being cancerous, where detection
CC of the differentially expressed gene product is correlated with a
CC cancerous state of the cell from which the test sample was derived.
CC The polynucleotide sequences can be used in a method for detecting
CC differentially expressed genes correlated with a cancerous state of a
CC mammalian cell. The polynucleotides can also be used as probes for
CC detecting and mapping related genes. They can be used in diagnosis and
CC prognosis of diseases and disorders (e.g. identification of
CC pre-metastatic or metastatic cancerous states, stages of cancer, or
CC responsiveness of cancer to therapy). This is particularly for breast
CC cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-
CC negative breast cancer, lung cancer, and colon cancer.
XX
SQ Sequence 733 BP; 218 A; 145 C; 152 G; 210 T; 8 other;

Query Match 10.4%; Score 656.8; DB 21; Length 733;
Best Local Similarity 97.5%; Pred. No. 7e-169;
Matches 696; Conservative 0; Mismatches 13; Indels 5; Gaps 3;

Qy 1405 catatgaccacgtcatatcacagcctcaaaagagctcttgactgtgagctgcagagccc 1464
Dbb 20 cagcagaaccaagtcatacacagcctcaaaagagctcttgactgtgagctgcagagccc 79

Qy 1465 agttgcataccactgccattgacaaagagggtcgtcggtgttaaaagcgggagctatg 1524
Dbb 80 agtgccttcactgcctgacaaagagggtcgtcggtgttaaaagcgggagctatg 139

Qy 1525 ctgcttgccaggaagcaaaagagataataaagagtcattcagaaaaatgtctccaacatc 1584
Dbb 140 ctgcttgccaggaagcaaaagaggattttaaagagtcattcagaaaaatgtctccaacatc 199

Qy 1585 cacttcattgagaagtattacactcagagattatgctcattcagaaaaattgctttgcgc 1644
Dbb 200 cacttcattgagaagtattacactcagagattatgctcattcagaaaaattgctttgcgc 259

Qy 1645 ttggttcctggatgaacaaaattatagctattcaagtgagcttagcagatctttgccc 1704
Dbb 260 ttgg-tcctggatgaacaaaattatagctattcaagtgagcttagcagatctttgccc 318

Qy 1705 aagcatgccttagagaagaacacctgactcggagaaatccctctcattcagaggttaatgc 1764
Dbb 319 aagcatgccttagagaagaacacctgactcggagaaatccctctcattcagaggttaatgc 378

Qy 1765 ttgggatgcaaaagctttataaaaggtgccgttaagatcccttcattgaattgatcttcagca 1824
Dbb 379 ttgggatgcaaaagctttataaaaggtgccgttaagatcccttcattgaattgatcttcagca 438

Qy 1825 gttttttttatgagatggaaatacaaaaaactcttgcctatggaatttgcgaagtattata 1884
Dbb 439 gttttttttatgagatggaaatacaaaaaactcttgcctatggaatttgcgaagtattata 498

Qy 1885 acaactgcagaagaataatatacagtgatcatcagacagagtatctctataactcac 1944
Dbb 499 acaactgcagaagaataatatacagtgatcatcagacagagtatctctataactcac 558

Qy 1945 ttccagttcagatgtttactgttcctactcctgctgcagactcttatttgaagagcagaatg 2004
Dbb 559 ttccagttcagatgtttactgttcctactcctgctgcagactcttatttgaagagcagaatg 618

Qy 2005 ttatctctgctactgaactctgtagaagttttacacctgagcttagctggacagg--aa 2062
Dbb 619 ttatctctgctactgaactctgtagaagttttacacctgagcttagctggacagg--aa 678

Qy 2063 caataaattcaacttccagggttatagcc--aggacaaattgggaagagtatat 2114
Dbb 679 caataaattcaacttccagggttatagccaggcagggaacaaattgggaagagtatat 732

RESULT 9

AAH07621	1255	ttgtccaagccagagaaataatttcttcagtgataaaataatgtcgtgagaatgactatat	1314
ID	AAH07621	standard; cDNA; 712 BP.	
XX			
XX	68	ttgtccaagccagagaaataatttcttcagtgataaaataatgtcgtgagaatgactatat	127
AC			
AAH07621;			
26-JUN-2001	(first entry)		
XX			
DE	Human cDNA clone (5'-primer)	SEQ ID NO:4456.	
XX			
KW	Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.		
XX			
OS	Homo sapiens.		
XX			
PN	EP1074617-A2.		
XX			
PD	07-FEB-2001.		
XX			
XX	28-JUL-2000; 2000EP-0116126.		
PF			
XX	29-JUL-1999; 99JP-0248036.		
PR	27-AUG-1999; 99JP-0300253.		
PR	11-JAN-2000; 2000JP-0118776.		
PR	02-MAY-2000; 2000JP-0183767.		
PR	09-JUN-2000; 2000JP-0241899.		
XX			
PA	(HELI-) HELIX RES INST.		
XX			
PI	Ota T, Isoqai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;		
PI	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;		
XX			
DR	WPI; 2001-318749/34.		
XX			
PT	Primer sets for synthesizing polynucleotides, particularly the 5602		
PT	full-length cDNAs defined in the specification, and for the detection		
PT	and/or diagnosis of the abnormality of the proteins encoded by the		
PT	full-length cDNAs -		
XX			
PS	Claim 1; SEQ ID 4456; 2537pp + CD ROM; English.		
XX			
CC	The present invention describes primer sets for synthesising 5602		
CC	full-length cDNAs defined in the specification. Where a primer set		
CC	comprises: (a) an oligo-dT primer and an oligonucleotide complementary		
CC	to the complementary strand of a polynucleotide which comprises one of		
CC	the 5602 nucleotide sequences defined in the specification, where the		
CC	oligonucleotide comprises at least 15 nucleotides; or (b) a combination		
CC	of an oligonucleotide comprising a sequence complementary to the		
CC	complementary strand of a polynucleotide which comprises a 5'-end		
CC	sequence and an oligonucleotide comprising a sequence complementary to a		
CC	polynucleotide which comprises a 3'-end sequence, where the		
CC	oligonucleotide comprises at least 15 nucleotides and the combination of		
CC	the 5'-end sequence/3'-end sequence is selected from those defined in		
CC	the specification. The primer sets can be used in antisense therapy and		
CC	in gene therapy. The primers are useful for synthesising polynucleotides,		
CC	particularly full-length cDNAs. The primers are also useful for the		
CC	detection and/or diagnosis of the abnormality of the proteins encoded by		
CC	the full-length cDNAs. The primers allow obtaining of the full-length		
CC	cDNAs easily without any specialised methods. AAH03166 to AAH13628 and		
CC	AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to		
CC	AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632		
CC	represent oligonucleotides, all of which are used in the exemplification		
CC	of the present invention.		
XX			
SQ	Sequence 712 BP; 229 A; 132 C; 156 G; 190 T; 5 other;		
Query Match	10.3%; Score 647; DB 22; Length 712;		
Best Local Similarity	97.6%; Pred. No. 3.3e-166;		
Matches 687; Conservative	0; Mismatches 13; Indels 4; Gaps 3;		
QY	1195	aacctggaagcagcagctactataaaagaattcacgctgtccgttgaaatgaagaggttaa	1254
Db	1		
	8	acctggaagcagcagctactataaaagaattcacgctgtccgttgaaatgaagaggttaa	67

CC The present sequence represents a human expressed sequence tag (EST).
CC The polynucleotide, which is a secreted EST, and the encoded protein
CC are predicted to have useful biological activities which would make
CC them suitable for treating, preventing or ameliorating medical
CC conditions in humans and animals, although no supporting data is
CC given. Suggested activities include nutritional activity, immune
CC stimulating or suppressing activity, haematopoiesis regulating
CC activity, tissue growth activity, activin/inhibin activity,
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic
CC activity, receptor/ligand activity, anti-inflammatory activity,
CC cadherin/tumour invasion suppressor activity, tumour inhibition
CC activity. The polynucleotide may also be useful for gene therapy.
XX
SQ Sequence 455 Bp; 108 A; 121 C; 107 G; 119 T; 0 other;

	Query Match	2.5%;	Score 159;	DB 20;	Length 455;
	Best Local Similarity	61.9%;	Pred. No. 4.3e-33;		
	Matches 252;	Conservative	0; Mismatches 155;	Indels	0; Gaps
Qy	5529	agtgccacgctcgcagatgatgacgaagaatcctcttcccttcacctttctgtggg	5588		
Db	451	AGGTGCCCAATCAGTGGTGATTAGACGAGCCCCAACTCGTCCTTGTCGCGGA	392		
Qy	5589	gctatactatgtctcagaaactattgctgccaggaaaattggaacggggaagagtgtgga	5648		
Db	391	TCTCTGCTGTGCTCCAGAGTTACTGCTGCCAGACTGAAC TGGAAGGGGAGGATGTAGGA	332		
Qy	5649	gcttgcatatttcacgacacttcactgtggagccggagctcgacttttccctaaaaaatcaga	5708		
Db	331	GCTGCACAGCTCACACCTPACCTCTGGCTCTGGAGTGGGCATCTTCTCGAGAGTAGCG	272		
Qy	5709	gaatgccgagtggtcctgggtgaaggtaaagccagagctgtgctatccagctcccttac	5768		
Db	271	GAATGTCAGGTGCTATTTTAC TCGSAAAAACAAGGCTGTTTTATCTCCTCCCTTAC	212		
Qy	5769	tggagtgaatatggagaaacagaccctcgctgtaagaggggcaaccccccttatctatct	5828		
Db	211	CTTGATGATTATGGGAGACCAGCGGSACTTCAGACGGGAAAATCTCTTACATTTATGC	152		
Qy	5829	cftgagcggatcgaagagctccatttgtctctggcaacaacactgcattatagaagagatt	5888		
Db	151	AAGACCGATTCAAGNAGATT CAGAAGCTCTGGCACCAACACAGTGTCTACAGAGAAATT	92		
Qy	5889	gctaggagccaagactaatcacagtgtatttggattccaactggca	5935		
Db	91	GGACATGCACAGGAAGCCAATCAGACACTGTTGGCAITGGACTGGCA	45		

RESULT	15	
AAF58252/C		
ID	AAF58252	standard; DNA; 936 BP.
XX	XX	
XX	AC	
XX	AAF58252;	
XX	24-APR-2001	(first entry)
DT		
XX	Oligonucleotide	D1835.
DE		
XX	Electron-transfer group;	ETM; mismatch; genotyping;
KW	gene expression; ss.	
XX	Synthetic.	
OS		
XX	W0200107665-A2.	
PN		
XX	01-FEB-2001.	
PD		
XX	26-JUL-2000; 2000WO-US20476.	
PF		
XX	26-JUL-1999; 99US-0145695.	
XX	17-MAR-2000; 2000US-0190259.	
PR		
XX		

PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
XX Umek RM;
XX
XX WPI; 2001-159728/16.
XX
XX Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
XX Example 6; Page 127; 159pp; English.
PS
XX The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
XX Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;
SQ

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 1, 2002, 12:39:35 ; Search time 4409.56 Seconds
(without alignments)
23599.681 Million cell updates/sec

Title: US-09-724-126A-1
Perfect score: 6308
Sequence: 1 gccagaattgcgcagagg.....aatttgattgtgtgtttt 6308

Scoring table: IDENTITY_NUC
Gapop 10_0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues
Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl : *
1: gb_ba : *
2: gb_hgt : *
3: gb_in : *
4: gb_cm : *
5: gb_ov : *
6: gb_pat : *
7: gb_ph : *
8: gb_pl : *
9: gb_pr : *
10: gb_ro : *
11: gb_sts : *
12: gb_sy : *
13: gb_un : *
14: gb_vi : *
15: em_ba : *
16: em_fun : *
17: em_hum : *
18: em_in : *
19: em_om : *
20: em_or : *
21: em_ov : *
22: em_pat : *
23: em_ph : *
24: em_pl : *
25: em_ro : *
26: em_sts : *
27: em_sy : *
28: em_un : *
29: em_vi : *
30: em_htgo_hum : *
31: em_htgo_inv : *
32: em_htgo_rod : *
33: em_htg_hum : *
34: em_htg_inv : *
35: em_htg_rod : *
36: em_htg_other : *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	4435.4	70.3	6395	6	AR030784	AR030784 Sequence
2	4435.4	70.3	6395	6	AR121463	AR121463 Sequence
3	4435.4	70.3	6395	10	AF061555	AF061555 Mus muscu
4	2546.8	40.4	2550	9	AK027803	AK027803 Homo sapi
5	996.2	15.8	1001	6	AR030785	AR030785 Sequence
6	996.2	15.8	1001	6	AR121464	AR121464 Sequence
7	994.2	15.8	999	9	AF061556	AF061556 Homo sapi
8	981	15.6	6158	9	AB002347	AB002347 Human mRN
9	764.2	12.1	818	9	HSX521	247040 Human part1
10	651.6	10.3	2958	9	AK026998	AK026998 Homo sapi
11	503.8	8.0	166518	9	AC090514	AC090514 Homo sapi
12	487.8	7.7	2512	9	AK026948	AK026948 Homo sapi
13	354	5.6	190727	9	AC068724	AC068724 Homo sapi
14	279.4	4.4	61896	2	AC090992	AC090992 Homo sapi
15	215.2	3.4	910	10	MM1UBR2	AF067372 Mus muscu
16	212.2	3.4	107304	2	AC016274	AC016274 Homo sapi
17	212.2	3.4	164468	2	AC021899	AC021899 Homo sapi
18	180.8	2.9	164468	2	AC021899	AC021899 Homo sapi
19	160.6	2.5	3980	9	HS1UBR4	AF067383 Homo sapi
20	150.2	2.4	107304	2	AC016274	AC016274 Homo sapi
21	132	2.1	28834	2	AC017982	AC017982 Drosophil
22	132	2.1	152545	3	AC010922	AC010922 Drosophil
23	132	2.1	179016	3	AC018489	AC018489 Drosophil
24	130.4	2.1	900	9	HS1UBR3	AF067382 Homo sapi
25	123.4	2.0	907	10	MM1UBR1	AF067371 Mus muscu
26	115.2	1.8	61896	2	AC090992	AC090992 Homo sapi
27	111.8	1.8	2616	10	MM1UBR6	AF067376 Mus muscu
28	106.4	1.7	830	9	HS1UBR5	AF067384 Homo sapi
29	104	1.6	143291	9	AL137792	AL137792 Human DNA
30	104	1.6	143697	2	AL358572	AL358572 Homo sapi
31	104	1.6	295265	2	AL353354	AL353354 Homo sapi
32	102.6	1.6	716	10	MM1UBR5	AF067375 Mus muscu
33	101.8	1.6	1000	9	HS1UBR6	AF067385 Homo sapi
34	100.8	1.6	1370	9	HUMORF02	D14658 Human mRNA
35	100.4	1.6	1100	10	MM1UBR4	AF067374 Mus muscu
36	97	1.5	179588	2	AC009852	AC009852 Homo sapi
37	97	1.5	182660	2	AC012416	AC012416 Homo sapi
38	97	1.5	183982	2	AC009825	AC009825 Homo sapi
39	91.4	1.4	1000	9	HS1UBR7	AF067386 Homo sapi
40	91.2	1.4	727	10	MM1UBR7	AF067377 Mus muscu
41	89	1.4	789	4	CFU12687	U12687 Canis famil
42	88	1.4	168639	2	AP001992	AP001992 Homo sapi
43	88	1.4	213613	2	AC013405	AC013405 Homo sapi
44	86.4	1.4	400	9	HS1UBR1	AF067380 Homo sapi
45	86	1.4	195064	2	AL138921	AL138921 Homo sapi

ALIGNMENTS

RESULT 1
AR030784
LOCUS AR030784 6395 bp DNA
DEFINITION Sequence 1 from patent US 5861312.
ACCESSION AR030784
VERSION AR030784.1 GI:5943998
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 6395)
AUTHORS Varshavsky, A. and Kwon, Y. Tae.
TITLE Nucleic acid encoding mammalian UBR1
JOURNAL Patent: US 5861312-A 1 19-JAN-1999;
FEATURES Location/Qualifiers
source 1..6395
/organism="unknown"
BASE COUNT 1802 a 1376 c 1519 g 1698 t
ORIGIN

PAT 29-SEP-1999

Query Match	70.38;	Score 4435.4;	DB 6;	Length 6395;
Best Local Similarity	86.3%;	Pred. No. 0;		
Matches 4997;	Conservative 0;	Mismatches 721;	Indels 69;	Gaps
QY	590	tccaggggccgtcgtataaagtgcgtccctcgtctc-tccgacggccacaggtttccgct	648	
DB	8	TTcAGGGCCGTCGTAAAGTGTCTCCCTGTCTGGTCTGGCGCCGCACAGGTTTCGGCT	67	
QY	649	tgcctctggccgggggtcgccgaactgcaggcgctcagtttccctcaagatgcccagagagag	708	
DB	68	AGCTGGCGCGCGGGTCTGGGAACTGGCGGGTTCGTTTCCCTTAAGTAGCGGACGAAG	127	
QY	709	aggctgaaggtactgagaggatgaaatcagcgcggagttacccccagagacccctcagcgtc	768	
DB	128	AGATGGACGGCGCCGAGAGGATGACGTCAGCCGGAGCCTCCCTTGGCCCCGACGGCG	187	
QY	769	tggcatcttgtgggtagcagcaagtgtatttttatactgctttcttgatcatctttggcac	828	
DB	188	CGGCATCTGCTGGGATCAGCAAGTGTATCTATATCTGCTTCTTACATCATTTGGCAC	247	
QY	829	aattgggtccagaaatttaacttctgtgaaatggaccacagacttgaaaaagcaggagaaa	888	
DB	248	AATTAGTCCCAAGAAATTTATTTTCTGATGGATGGACCCAGATTTGGAAAACAAGAGA	307	
QY	889	gtgtacaaatgcaatttactccactccactgaaatgtagttatttttgagagatccagata	948	
DB	308	GTGTACAGATGTCAATATCTACTCTTTGGAGTGGTACTTATTTGGAGAGATCCGGATA	367	
QY	949	tttgcttagaanaattgaagcacagtgagcatttcagcttctgggaggggttttcaaaa	1008	
DB	368	TTTGCTTAGAGAAATTAACACACAGTGGAGCGTTCAGTGTGTGGGAAGGTTTCAAAA	427	
QY	1009	gtgagagagaaacctattctctgcagggattgtgcaattgatacacaatgctgactctgta	1068	
DB	428	GTGAGAAACAACATATTCCTGTAGGATTTGTGCAATTTGATCACAACGTGTGCTCTGTA	487	
QY	1069	tggactgtcttcaggagacagtggtcataaaaatcatcgttacagatgcatcttactctg	1128	
DB	488	TGGACTGCTTCCAAAGTAGTGTTCATAAAAACCATCGTTACAAGATGCATCTTCTACTG	547	
QY	1129	gagaggggtctgtgactgtggagacacagggcattggaaaactggcccttttttgtata	1188	
DB	548	GAGGGGGCTTCTGTGACTGTGGACACACAGAAGCGTGGAAAACTTGGCCCTTTTGTGTGG	607	
QY	1189	atcatgaacctggagagcagggtactataaaagagaattcacgctgtccgttgaatgaag	1248	
DB	608	ATCAGAGCCTTGGGAAGACAGGTTACTACAAAGAGAGCTTTACATTGCCCATTTGAATGAAG	667	
QY	1249	aggttaattgtccagccaggaataattctctcagtgataaaatgtcgtagaataga	1308	
DB	668	AGGTGATTGCTCAGCCAGGAGAAATATTCCTTCGTGTATAAATACATTGTAGAAATGA	727	
QY	1309	ctatatgggaagagaaaagaactgcctctgaactccagataaaggagaaaaatgaaa	1368	
DB	728	CTATATGGGAAGAGAAAGAAATTCCTCTCTGAACATGCAGATAAGGAGAAAAATGAAC	787	
QY	1369	gatactatgtgctctttcaatgatgaacacaaatcatatgaccagtcataacagcc	1428	
DB	788	GATACTATTGTGTCTTTTCAACCATGAGCACCAATTCGTATGATCATGTGATCTACAGTC	847	
QY	1429	tacaaaagctcttgactgtgagctgcagagccaggttgataccactgcccattgaca	1488	
DB	848	TGCAGAGAGCTTATAGTTGCAGGCTTGACAGGGCACAGCTTGCACACAGCTGCCATCGACA	907	
QY	1489	aagaggggtcgtcgggctgttaaacggggagctlatgctgcttggccaggaagcaaaaggaag	1548	
DB	908	AAGAGGTCGCGGGCTGTCAACACAGGTGTGTATGCCACTTCCAGGAAGCAAGGAGG	967	
QY	1549	atataaagagtcaatcagaaatgtctctcaaacatccacttcattgtagaagtattacact	1608	
DB	968	ATATAAAGAGTCACTCAGAGAAGCTGTCTCAGACACCCCTTCCATGTGTGAAGTGTCTGCAT	1027	

Qy	1609	cagagattatggctcatcagaaatttgcttggcttggctcctggatgaacaaatta	1668
Qy	1610	tt	1669
Db	1028	CCGTGGTATGTGCTACCAAAATTCGCTTGCGCCTTGCTCCTGGATGAACAAATTA	1087
Qy	1669	tgagctattcaagtgaacttaggcagatcttttgcgaagcatgccttagagaagaacctg	1728
Db	1088	TCAGCTATTCAAGTGACTTTAGACAGATATTTTGGCCAGGCTGCCTGTAGAAGAACCCTG	1147
Qy	1729	actcggagaaatccccctgtctcataagcagggttaatgctttggggatgcaaaagctttataaag	1788
Db	1148	GCTCTGAAAATCCCTGCTTATAAGCAGACTAATGCTTTGGGATGCAAAACTTTATAAAG	1207
Qy	1789	gtgcccgtaagatccttcgatgaattgatcttcagcagttttttttaggaatggaataaca	1848
Db	1208	GTGCCCGTAAGATCCTTCATGAATTCATCTTTTASTAGTTTTTTTTTATGGAGATGSAATACA	1267
Qy	1849	aaaaactcttgcattggaatttggaagtcattataaaacaactgcagaagaatatatca	1908
Db	1268	AAAAACTCTTTGCTATGGAAATTTGTGAAGTATTTATTAACAACTGCAAAAGAGATCATCA	1327
Qy	1909	gtgatgatcatgacagaagtatactctataactgcactttcagttcagatgtttacttgttc	1968
Db	1328	GGCAGCACACGAGAGAAGCATCTCCATTAACGCCCTGTCCGTGCAGATGCTCACCGTCC	1387
Qy	1969	ctactctggctgcagacatttattgaagacagaaatgttatctctgtcattactgaaactc	2028
Db	1388	CGACCTTGGCCCGCATCTTATTGAAGACGAGAACTGTTATTTCTGTCTATTACTTGAACGC	1447
Qy	2029	tgttagaagttttacctgagttacttgacaggaacaataaaatccaacttccagggttata	2088
Db	1448	TGCTAGAAGTTTTTACCCTGAATACTTGGACAGGAACAATAAATTTCAACTTCCAGGGTTATA	1507
Qy	2089	gccaggacaaattgggaagagtatacgagtaatatgtgacctaaagtatatcctgatca	2148
Db	1508	GCCAGGACAACTGGGAAGAGTCTACGCAGTTATATGTGACCTTAAAGTATATCTCTGATTA	1567
Qy	2149	gcaaacccacaatatggacagaaagaattaagaatgcagttccttgaagtttttcgactctt	2208
Db	1568	GCAAGCCGTTCATATTTGGACAGAAACGATTAAGAGCGCAGTTCCTTGGAAAGGTTTTCCGGTCTT	1627
Qy	2209	ttttagaagattctacctgtatgcagggaatggaagaataatcgaagaacaggtttggacaac	2268
Db	1628	TTTCTGAAGATCTTACCTGTATGACGGGAATGGAAGAAATCAGAAGCAAAAGTTTGGACAAC	1687
Qy	2269	acattgaagtggatcctgtattgggaggctgccattgctatacagatgcgaattgaagaata	2328
Db	1688	ACATTTGAAGTGGACCCCTGACTGGGAGGCTGCCATCGCTATACAGATGCAACTAAAGCAATA	1747
Qy	2329	ttttactcatgttccaaagatgggtgtgtgtgtgtatgaaagaactcttacttgtggtctata	2388
Db	1748	TTTTTGCTCATGTTCCAAGAGTGGTGTGCTTGTGTGATGAAGATCTCTTACTGGTGGCTTATA	1807
Qy	2389	aagaatgtcacaaagctgtgatgaggtgcagttaccagtttcatctcagtagcaagaacag	2448
Db	1808	AGAATGTACAAAGCTGTAAATGAGTGGCAGTACAAATTTTCATGTCCAGTACCAGACAG	1867
Qy	2449	tagtacaatcgttggacatagtttggaaacaaagtccctcacagatcatctgagagactcttg	2508
Db	1868	TAGTCCAAATTTGGCGGTCTCATAGTCTGGAACCAAACTCTTACAAAGTGTCTGAGGACCTTG	1927
Qy	2509	taagcatatcatctgcactctctaggaaccttgcgtggtcttcattgtacgttttaagcagtc	2568
Db	1928	TAAAGCATACACCTGCCACTCTCTAGAACAACCTTGTCTGTCATGTACGCTTTAAGCAGAC	1987
Qy	2569	tgggtgctgtttcaagactgcatgaatttgttcttttgagacatttcaagtagaggtac	2628
Db	1988	TAGTGTCTATTTCAGACTGTCATGAATTTTGGCTTTTTCAGAGCTTTCAAGTAGAGGTCC	2047
Qy	2629	tagtggaaatacctttacgttctcgt	2688
Db	2048	TGGTGGAGTAGCCCGCTGGCTGCCCTGGTCTGCTGCTCAGGTGTGTGTCTGAGATGTGGC	2107
Qy	2689	gaagaatatggactgtctcttattagccaggtgtttttattaccaagaatgtttaagtcagag	2748

Db 2108 GAAGAAACGGGCTCTCACTCATCAGCAGGTTTCTATTATCAAGATGTTAAATGCAGGG 2167
Qy 2749 aagaaatgatataaagatatcatcgtcttcagattggtgcaltcttaagtatccca 2808
Db 2168 AGGAAATGACGATAAAGATATCATCTGCTTCAGATTGGAGCAATATATAGATCCCA 2227
Qy 2809 ataagttcttggtacttcagaggtatgaacttgcagagcttttaacaagacca 2868
Db 2228 ACAAGTTCTTGTTACTTGTTACTTCAGAGATATGAACCTTACTGATGCTTTTAAACAGACCA 2287
Qy 2869 tatctacaaagaccagatttgattgtaataacaataataacactaatagaagaatgcttc 2928
Db 2288 TATCCAAAAAGACACGAGTTTGATTAACACAGTATAATACATTAATAGAGAAATGCTTC 2347
Qy 2929 aggtccctcatctatttggtgagcgttatgtacctggaagtggaaatgtaccaaaag 2988
Db 2348 AGGTCCTCATCTATTGTGGGAGACGTTATGTACCTGGAGTGGGAAATGTTACCCAGAG 2407
Qy 2989 aagaggtcaaatgagagaatcattcacttgccttgattgaaccccatgccacacagt 3048
Db 2408 AGGAGGTTATAATCAGAGAGATTACTCATTACTTTCATTGCGATGCCACACAGTG 2467
Qy 3049 ccattgccccaaatttaccctgagaatgaaataatgaaactggcttagagaatglcataa 3108
Db 2468 CCATCGCCACAAAACCTACCTCAGAACGAAAAATAATGAACCTGGCTTAGAGAAATGTCATAA 2527
Qy 3109 acaaagtgccacatttaagaacaaagtgatcagcccatggatttatgaactaaaag 3168
Db 2528 ACAAGTGGCCACAAATTAAGAAACACGAGTGTGCGGCCATGGAGTTTATGAATTGAAG 2587
Qy 3169 atgaactgaagaagactcaaatgtactttatctattactcacaacccagcatagca 3228
Db 2588 ATGAATCACTGAAGACTTCAATATGTACTTTTACCATTATCTTAARACACAGCATAGCA 2647
Qy 3229 aggtgaacatattgcagaagaaggagaaacaaagaagaaacaaagatgaagcattgcgc 3288
Db 2648 AGGCTGAACATATGCAGAAGAAAGAGAAACAAAGAAATAAAGATGAAGCATTGCCGC 2707
Qy 3289 caccacacctctgaattctgcctctgttcagcaagtgattaaacctctcaactgtg 3348
Db 2708 CGCCACCTCTCCAGAGTTCTGCCCTCTTTCAGCAAGTAGTCAACCTGTGCTAGCTGTG 2767
Qy 3349 atatcatgatcatctcagaccgtatttgagcgggcaatagacacagattcact 3408
Db 2768 ATGTTATGATATACATCTCAGGACCAATCTTGGACGGGCAGTGGACACGAGTCTAATC 2827
Qy 3409 tgtgaccgaaggatgtctccaaatggcttttcattctggcattgggtttactagaag 3468
Db 2828 TGTGGACAGAAGGATGCTGCAGATGGCTTCCATATATTGGCAGCTGGGCTTGTGGAAG 2887
Qy 3469 agaagcaacagcttcaaaaagctcctgaagaagaagtaacatttgacttttcatcaagg 3528
Db 2888 AGAAGCAGCAGCTTCAGAAAGCTCTCAGAGGAGTCTTGAAGAGGAGTGGCTTTTACCATAAG 2947
Qy 3529 ctcaagattgggaagttcagccat-----gaatatacaaatgcttttgaaaaac 3579
Db 2948 CTTCAAGATTGGGAGTTACCCATGAATGCTCAGATATACAAATGCTCTTGGAAAGAC 3007
Qy 3580 tcaaggaattccccagtttagaaggccagagggacatgataacgttgatacttcagatgt 3639
Db 3008 TCAAAGGAATCCCCAATTAGAAGGCCAGAGGACATGATAACATGGATACTCCAGATGT 3067
Qy 3640 ttgacacagtgagcgttaagagaaaaattctgtttaattgtagcaaccacatcaggat 3699
Db 3068 TTGACACAGTGAAGCGATTGAAGAAAAATCTGTTTAGTTGTGGCAACCACTTCAGGAC 3127
Qy 3700 cggaaatctattaagaatgatgagatttactcatgataaagaaaaagcagaacgaaaaagaa 3759
Db 3128 TGGAGTGCAATTAGAGTGAAGGATTAATCTCATGATAAAGAAAAAGGCAGACGGGAAGAA 3187
Qy 3760 aagctgaagctgttaggtcatctgccagaagatctatggctcagatgtctgctctcaga 3819
Db 3760 aagctgaagctgttaggtcatctgccagaagatctatggctcagatgtctgctctcaga 3819

Db 3188 AAGCTGAGGCCCTAGGCTTCATGCGCCAGAGATCATGCGCCAGATGTGTGCTTTACAGA 3247
Qy 3820 aaaaactcattgaactcataaaactcatgtatgaacaatacatcagaataatgcctgggaaag 3879
Db 3248 AAAACTTCATTGAACCCACAACACTCATGTATGATAATACGTCAGAACTTAACAGGAAG 3307
Qy 3880 aagattccattatgagggaagagacccccagctcagtgactactctagaattgctt 3939
Db 3308 AAGACTCCATTATGGAGGAAGAGAGCACCTCAGCAGTCAGTGAGGCTCTAGAAATGCTC 3367
Qy 3940 tgggtccctaaacggggtccactgttactgaaagagggtgctgacgtgcacatcttggcc 3999
Db 3368 TGGGCCCTAAACGGGCCCGGCTGTGTACCGAAAAGAGGTGCTGACGTGCATCTCTGCC 3427
Qy 4000 aagaagaacagagatgaaaaatagaaaaataatgccatgtattatcgcgctgtgtccaga 4059
Db 3428 AAGAAGAACAGAGGTAAACCTAGAAAAATAATGCCATGTTTGTGAGCATGTGTGCAGA 3487
Qy 4060 aatctactgccttaaccccgagcaggggaaaccccatagaactctcaggagaagccctag 4119
Db 3488 AATCCACGCCCTAACCCAGCACAGAGGGAAGCCTGTGCACCACCTTAGGGGAACACTGG 3547
Qy 4120 accacttttcatggatccagacttggcataatggaacttatatacagaagaagctgtggtcag 4179
Db 3548 ACCCTCTTTTTCATGGATCCAGACTTTGGCACATGGAACTTTATACAGGAAGCTGTGTCATG 3607
Qy 4180 taatcacagcagctgtctgacagaagtatttgaagctgtacagctcagctcagctcagcagc 4239
Db 3608 TAATGCATGCAGTGTCTGGCAGAGATATTTTGAAGCTGTGCAGCTGAGCTCCGACAGC 3667
Qy 4240 gcattcatgttgaccttttggacttggaaagtggaaataatcttttccctcttttggcaat 4299
Db 3668 GCATTCACGTAGACCTGTTTGACCTGGAGAGCGGCGAGTACCTATGCCCCCTCTGCAAGT 3727
Qy 4300 ctctgtgcaatactgtgatcccaattattcccttggcaacctcaaaagataaacaagtgaga 4359
Db 3728 CTCTCTGCAACACTGTCTACCTCCCATCATCCCTTTGAGCGCGGAGAGATCAACAGTGAGA 3787
Qy 4360 atgcagatgctcttgctcaacttttgaccttggcagcgtggacagtgagatacagactgttgcgca 4419
Db 3788 ATGCGGAGGCTCTTGCTCAACTTTTGACCTTGGCCGGTGGATACAGACTGTCTTGCCA 3847
Qy 4420 gaatacaggttataataaagacatgctaaaggagaaacac-----caattccctattttct 4476
Db 3848 GAATATCGGTTTATATATAAGCATGCTTAAGGAGAAAGCCCGCAGCAGTTCCTGTCTGT 3907
Qy 4477 ttaataaagggaatggagatctctactttggagtccactccatccctgagtttggcgttg 4536
Db 3908 TTAATCAAGGAATGGGGATTCAACTTTTGAAGTTTCAATCCATCTGAGTTTTGGAGTTC 3967
Qy 4537 agcttcgattaaataattcaaatagcatcaagaaaatggttattctcttcttgcacaaaca 4596
Db 3968 AGTCTCGGTGAAATATTCAAATAGTATCAAGGAATGTCATTCTCTTCGCCACAAACAA 4027
Qy 4597 ttatagattggattgaaagtgcacatgtgaaaggatcctcagtcagtcctccatgctga 4656
Db 4028 TTTACAGAAATGGCCTGGAAGTGCCTCCCTGATGAAGTACAGCCACGAGTGCCTCATGATGA 4087
Qy 4657 cctggagcacctgcgcttccactatccaggaattgaaatctattgggagatgaaggaa 4716
Db 4088 CCGGAGCAGCGTGTGCGTTCACTCAGCAATCCAGAAACCTGTGGGAGATGAAGGAA 4147
Qy 4717 aacctctgtttggagcaacttcaaaaataggcagcataatgtgtgaaagcattatcagct 4776
Db 4148 AACCTCTATTTCGAGCAGCTTCAAAATAGACATAGCGGTCTCTGAAGGGGCTAATGCAGT 4207
Qy 4777 ttgcattgacagaggtattacctgtcctcaggtcctgatagacagaacaatctgttctgctc 4836
Db 4208 TTGCAGTTGCACAGAGGGCTACTTGCCTCAGTCTCTGATACACAAACATCTCGCTCGGC 4267
Qy 4837 ttctatcagttgttcttccatacaataaatcagaagatacacacctccttctgtctatag 4896
Db 4268 TCCTGTCACTTATTCTTCTAACCCTGCATCAGAAATACACCAGGCGCTTCTGTCTGTGG 4327

Qy 4897 atctgttctatgttttgggtggtgctgtagcattccatctcttcttatttggatgacc 4956
Db 4328 ATCTCTTCCATGTCTGGTGGGCGAGCTTAGCGTTCCCATCTCTTATTGGGATGACA 4387
Qy 4957 ctgttgaatgcagcctcttccagttagttcttccataaccaccttattcttccatt 5016
Db 4388 CCGTGGATCTGCAGCGTCCGCTTACCTTCTCATATACCACTCTCTCTCTCCATC 4447
Qy 5017 tgatcacatggcacacatgcttcacatactacttaccagtagaca-----cag 5064
Db 4448 TGATCACATGGCGACATGCTTCAGATACTCTTACACAGATACAGATGCTGCTCCAG 4507
Qy 5065 gctacaccttctgcaggttccaagaagacagtgtaagaggctcatccgcattcttcttct 5124
Db 4508 GGGCGCGCTGTCTGAGGCTCAAGAGATAGTGAGGAGGCTCGCTGTGCATCTGCTTCT 4567
Qy 5125 ttgcagaaattctcaatatcacagtggtccattggtggtgatattccctggctgatt 5184
Db 4568 TTGTGGAGTGTGCGACACACAGCGGCTCTACTGGGTGGGCTCCGCGCTGGTACC 4627
Qy 5185 tgtgggtctcaactgaagaaatggcatcaccccttattctgctgctgcatgtttttcc 5244
Db 4628 TGTGGCTCTCCCTGAGAACCGGATCACCCCTTACCTCCGCTGTGCTGCATGCTTTTC 4687
Qy 5245 actatttacttgggtaactccgctgaggaactgcataccaaattctgcgaagagagt 5304
Db 4688 ACTATTTACTTGGAGTAGCTCCGCGCTCAAGAACTGTTTGCCTAATCTGCTGAAGGAGAT 4747
Qy 5305 acagtgcactctgtagctatctattcttaccatacaaatgttctctctcttccaggaat 5364
Db 4748 TCAGTGCACCTGTGTAGTATCTATCTTTACCCACAAATTTGTCTCTCTTTTCCAGGAAT 4807
Qy 5365 atlgggactatgaagcccttctccagaggtggtgtagatccctgccttactaaact 5424
Db 4808 ATTGGGATACCAAGGCCCTTACTACAGAGGTGCTGTGGAGATCCCTTACTCAAGT 4867
Qy 5425 gtttgaagcaaaaaaacacccgttggtcaggttacccctagaaaaaagaaatagtttgatagc 5484
Db 4868 CTTTGAAGCAGAAAAGTGTGTGTGTCAGGTACCTAGAAAAGAAATAGTTGTATAGAGC 4927
Qy 5485 ttctgtactatagctgctcctctgaatcaagcttctcttcaagtgccacaggtctg 5544
Db 4928 TTCTTGAGGACTACAGCTGCTTCTAAATCAGGCTTCTCACTTTAGTGTGTCACGGTCTG 4987
Qy 5545 cagatgatgagcaaacatccctgctctctctctctctctctctctctctctctctctc 5604
Db 4988 CAGATGATGAGCAAAACATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 5047
Qy 5605 agaactttgtccaggaaattgtgaacggggaagaggttggagcttgcaatttttcacg 5664
Db 5048 AGAACATCTCTTGCCAAAGAAATAGTGAATGGGAAGAGGTGGAGCGTGGCTTTTTCATG 5107
Qy 5665 caattcaactgagcggagctgctatcttcttaaaatacaagaatgcccagagtgctcc 5724
Db 5108 CGTTTCTTGTGTGTGAGTGTGATTTTCTTAAATCCGAGATGACAGGCTGTCC 5167
Qy 5725 tggatgaagtgaagcagaggtgctctatccagctcttacttgatgaatatgag 5784
Db 5168 TGGTGAAGGAAAAGCCAGAGGCTGTGCCATACCAGCCCTTACTTGGATGAATATGAG 5227
Qy 5785 aaacagaccttgctgaagagggggaaccccttctatttctctgtagcggtatcgga 5844
Db 5228 AACAGACCCAGGCTAAACAGAGGAAAACCACTTCTTATCTCGGAGCGGTATCGGA 5287
Qy 5845 agctccatttggctggcaacaacactgcatttatagagagattgctagagccaagaga 5904
Db 5288 AGCTGATTTGGCTGGCAACAGCACTGCATTTATAGAAAGATATGCTCGGAGCCAGAGA 5347
Qy 5905 ctaatcagatgtatttggattcaactggcagttactgtagctccaactctgcctcaag 5964
Db 5348 CTAATCAGATGCTATTGGAATTTAACTGGCAGTTTACTCTGAGCTTCAAGTTGCTCCTCAAG 5407

Qy 5965 acaatcacaaatgagcagtagtaaaaggctgattcaaaaattatggaaaaactttctgagg 6024
Db 5408 ACAATCATGAGTGACATCAATAATAAAGACTGATCTAAATTTCTAGAGAATTTCTGAGG 5467
Qy 6025 gctggaagatttggaggtcttttgcctatgctccaggttccactatcatcaataaaaat 6084
Db 5468 ACGGGGAAGTATTTGGAGGCTCTTTTGATCCATGTCAGATTCACACACATTAATAAAT 5527
Qy 6085 atttctaatggagtattgcttcttcaattagcaaacatatgcttccacagaaaaa-aggac 6143
Db 5528 ATTCCTTAATGCAATATTGCTTTCAATTATCAACATAAGCTTCAAGGGGAAAAACAAGAC 5587
Qy 6144 atagataaatctgt-----ttt 6160
Db 5588 ATAGATTAAATGTTTATGTTCTTAGAACACATAAAGAAATGCTTTTTCATCCAAAGTGCTAT 5647
Qy 6161 atgtctagtattccagaaatttattcccttcaataattgtctctattctatttattt 6220
Db 5648 TTCTGCTAATATTTCCAGAAAACCTCTTTCCCTTCATAACTGTCCTAGTTCATTTTCATAT 5707
Qy 6221 catccacttgtagaagtcacgtcaaacaggttgtagacatttttatgttgggttaac 6280
Db 5708 CACCACCTGGTTAATGAGGTCAATTAAGCATTTGTGACATTTTCCATCTGGCTAAC 5767
Qy 6281 tcttctgcaatttatttgggtttt 6307
Db 5768 ATCTGACCTTTGTTATTGTTGTTT 5794

RESULT 2
AR121463 LOCUS AR121463 6395 bp DNA PAT 16-MAY-2001
DEFINITION Sequence 1 from patent US 6159732.
ACCESSION AR121463
VERSION AR121463.1 GI:14105039
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 6395)
AUTHORS Varshavsky, A. and Kwon, Y. Tae.
TITLE Nucleic acid encoding mammalian Ubr1
JOURNAL Patent: US 6159732-A 1 12-DEC-2000;
FEATURES Location/Qualifiers
source 1. 6395
BASE COUNT 1802 a 1376 c 1519 g 1698 t
ORIGIN

Query Match 70.3%; Score 4435.4; DB 6; Length 6395;
Best Local Similarity 86.3%; Pred. No. 0;
Matches 4997; Conservative 0; Mismatches 721; Indels 69; Gaps 6;

Qy 590 ttcagggggcgtcgttaaaagtgtcctcctgtctc-tccgaccggccacaggtttccgct 648
Db 8 TTCAGGGGCGCTCGTAAAGTGTCTGCTCTGCTGCGGTGCGGCGCCACAGGTTCCGCT 67
Qy 649 tgcctctggcgggggttcggcaactgcaggcgtcagtttccctcaagatggcggacgag 708
Db 68 ASCTGGCGCGGGGTTCGGAACTGCGGGCGGTCTGCTTTCCCTTAAGATGGCGGACGAAG 127
Qy 709 aggtcgagggtactgagaggtatggaatcagcgcgaggtaccaccagaccctcagcgctc 768
Db 128 AGATGGACGGCCCGAGAGGATGGACGTACGCCCGGAGCCTCCCTGGCGCCGACGGCG 187
Qy 769 tggcatcttgggtggatcagcaagtgtgatttttatactgcttcttgcatactttggcac 828
Db 188 CGCATCTGTTGGGATCAGCAAGTTGATTCTATATCTTCTTACATCATATTTGGCAC 247
Qy 829 aattggtccagaaatttactttgctgtaaatggacccagacttgaaaagcggagggaaaa 888
Db 248 AATTAGTCCAGAAATTTATTTTGTGATGGACCCAGATTTGGAAGAACGAAGAGA 307

Qy 889 gtgtacaaatgtcaaatattcaactccactggaatggttacttatttttgggaagatccagata 948
Db 308 GTGTACAGATGTCAATACTCACTCCTTTGGAGTGGTACTTATTTGGAGAGATCCGGATA 367
Qy 949 tttgtttagagaaatgaagcacagtgagagcatttcagcttttgggaggttttcaaaa 1008
Db 368 TTTGCTTAGAANAATAAACAACAGTGGAGGTTCCAGTTGTGTGGGAAGGTTTTCAAAA 427
Qy 1009 gtggagagacaacctattcttgcagggatgtgcaattgatccaacatgtgtactctgta 1068
Db 428 GTGGAGAAACAACATATTCCTGTAGGGATTTGCAATGTATCCAAACGTGTGCTCTGTA 487
Qy 1069 tggactgttccaggagcagtggttccataaaaaatcatcgttacaagatgcatacttctactg 1128
Db 488 TGGACTGCTTCCAAAGTAGTGTTCATAAAAAACCATCTTACAAGATGCATACTTCTACTG 547
Qy 1129 gaggaggggtctgtgactgtggagacacagaggcatggaaaaactggccctttttgtgtaa 1188
Db 548 GAGGGGGCTTCTGTGACTGTGGAGACACAGAACGCTGGAAAACTGGCCCTTTTGTGTGG 607
Qy 1189 atcatgaacctgtgaagacaggttactataaaagagaattcacgcgtgccgttgaatgaag 1248
Db 608 ATCAGAGCCTGGAAGACGAGGTACTACAAAGAGAGCCTTACATTGCCCATTTGAATGAAG 667
Qy 1249 aggttaattgtccaggcaggaaaaatatttcttcaagtgataaaaaatagtctgtagaaatga 1308
Db 668 AGGTGATTGTCTCAAGCCAGCAGAGAATATTCCTTCGGTGATAAATACATTGTAGAAATGA 727
Qy 1309 ctatatgggaagggaagaaactgcctcctgaactccagataaaggagagaaatgaaa 1368
Db 728 CTATATGGGAAGAAAGAAAGAAATGTGCTCTGTAACCTGTCAGATTAAGGGAGAAAAATGAAC 787
Qy 1369 gatactatgtctcttttcaatgaatgaacacacattccatgaccacgtcatatataagcc 1428
Db 788 GATACTATTGTGCTTTTCAACGATGAGCACCATTTCGTATGATCATGTATCTACAGTC 847
Qy 1429 tacaaagactcttgcgtgtgactgcagagggcccagtttgcataccactgccaattgaca 1488
Db 848 TGCAGAGAGCTCTAGATTGGAGCTTGCAGAGGCACAGCTGCACACGACTGCCATCACA 907
Qy 1489 aagagggtctcgggctgttaaagcgggagcttatgcgtgttgcaggaagcaaggaag 1548
Db 908 AAGAGGTCGCGGGCTGTCAAAGCAGGTGTGTATGCCACTTGGCCAGGAAGCAAGGAGG 967
Qy 1549 atataaagatcattcagaaaaatgtctccaacatccaattccatgtagaagtattacact 1608
Db 968 ATATAAAGATCACTCAGAAACGTCTCTCAGCACCCCTCCATGTGGAAAGTGTGCACT 1027
Qy 1609 cagagattatggctcatcagaaaatttgcgttgggtcttgccttggaatgaacaaaatta 1668
Db 1028 CCGTGGTTATGGCTCACCAAAATTCGCTCTGCGCCTTGGCTCTCGATGAACAAAATTA 1087
Qy 1669 tgaacttaactgaactttagcagatcttcttggcgaagcatgccttagagaagaacctg 1728
Db 1088 TGAGCTATTCAAGTGACTTTAGACAGATATTTTGGCCAGGCTTGCCCTCTAGAACAACTG 1147
Qy 1729 actcgagaatccctgtctcataagcaggttaatgcttttgggatgaaagctttataaag 1788
Db 1148 GCTCTGAAAATCCGTGCTTATAGCAGACTAATGCTTTTGGGATGCAAAACTTTTATAAG 1207
Qy 1789 gtgcccgaagatcccttcagaaatgtatcttcagcagatttttttatggagatggaataca 1848
Db 1208 GTGCCGTAGATCCCTTCATGATTTGATCTTTAGTAGTATTTTATGAGATGGAATACA 1267
Qy 1849 aaaaacttttgcataggaaatttgcgaagtattataaacaactgcagaaagaatatatca 1908
Db 1268 AAAAAGCTTTTGTATGGAATTTGTGAAGTATTATAAACAACATGCACAAAAGAGTACATCA 1327
Qy 1909 gtgatgcatagcagaagtatctctataactgcactttcagtttcagatgcagatgttactgttc 1968
Db 1328 GCGACGACCAACGAGAGAAGATCTCCATAAACCGCCCTGTCCGTCAGATGCTCACCCCTCC 1387

Qy 1969 ctactcttggctcgacatcttattgaagagcagaaatgttatctctgtcattactgaaactc 2028
Db 1388 CGACCTTGGCCCGCATCTTATTATGAGAGCAGAAATGTTATTTGTCTCATTTACTGAAACGC 1447
Qy 2029 tgcataagatcttactcagtgacttggacaggaacataaaattcaacttccaggggtata 2088
Db 1448 TGCTAGAGATTTTACTCTGAATACTTGGACAGGAACAATAAATTTCAACTTCCAGGGTTATA 1507
Qy 2089 gccagagcaaaattgggaagatgatgtcagtaatatgtgacctaaagtatatctctgatca 2148
Db 1508 GCCAGACAAACTGGGAAGAGTCTACGCCAGTATATGTGNCCTAAAGTATATCTCTGATTA 1567
Qy 2149 gaaaaccacaaatatggacagaaaattgaagaatgcagtttcccttgaagggttttcgatctt 2208
Db 1568 GCAAGCCCTGCATATGGACAGAACGATTAAGAGCGCAGTTCTCTGGAAGGTTTCCGGTCTT 1627
Qy 2209 ttttgaagattcttaccctgtatgcaggggaatggaaagaaatccgaagacaggtttgggcaac 2268
Db 1628 TTTCTGAAGATCTTACCTGTATGTCAGGGAATGGAAGAAATCAGAAGACAAGTTTGGACAAC 1687
Qy 2269 acattgaagtgatcctgattgggaggtgcattgtctatcacagatgcgaattgaagaata 2328
Db 1688 ACATTGAAGTGGACCCTGACTGGGAGGCTGCCATCCTATACAGATGCCAACTAAAGATA 1747
Qy 2329 ttttactcatgttccaaagagtggtgtgtgtgatagaagaaactcttacttgtggcttata 2388
Db 1748 TTTTCTCATGTTTCCAAGAGTGTGTGTGTGATGAAGATCTCTTACTTGGTGGCTTATA 1807
Qy 2389 aagaatgtcaaaaagctgtgatgaggtgcagtcaccagtttcatatctagttagcaagacag 2448
Db 1808 AAGAAATGTCAAAAGCTGTAATGAGGTGCAGTACAAATTTTCATGTCCAGTACCAGACAG 1867
Qy 2449 tagtacaatcgtgtgacatagtttggaaacaaagtcctcacagatctctgaagatcttg 2508
Db 1868 TAGTCAATTTGTCGGTCTATAGTCTGGAAACCAAAATCTTACAAAGTGTCTGAGGACCTTG 1927
Qy 2509 taagcatcatctgcactctcttagacccttgcgtgtgttctaatgtacgttttaagcaggc 2568
Db 1928 TAAGCATACACCTGCCACTCTCTAGAACACTTGCCTGTCTTCATGTACGTTTAAGCAGAC 1987
Qy 2569 tgggtgcgtgttcaagactgcataaattgtgtcttcttttggaggacttccaagttagagttac 2628
Db 1988 TAGTGTCTATTTCAAAGACTGCATGAATTTGTGCTTTTGACAGCTTTTCAAGTAGAGTCC 2047
Qy 2629 tagtgaatactcttactgtgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2688
Db 2048 TGGTGGAGTACCCTGCGCTGCCCTGGTGTGCTGCTGAGGTGTGTGTGTGTGTGTGTGTGTGT 2107
Qy 2689 gaagaaatggactgtctcttattagccaggtgttttattaccagaagtgttaagtgcagag 2748
Db 2108 GAAGAAACGGGCTCTCACATCATCAGCCAGGTTTCTTATATCAAGATGTTAATGCGAGG 2167
Qy 2749 aagaatgtatgaataagataatcatctcttcaagatttggtgcattctttaaaggatccca 2808
Db 2168 AGGAAATGTACGATAAAGATATCATCATGCTTCAGATTTGGAGCATCTATAATGGATCCCA 2227
Qy 2809 ataaqcttcttactgttacttcagagatgaacttgcagaggtttttaaacaaccca 2868
Db 2228 ACAAGTCTTGTGTGTTACTGGTACTTCAGAGATATGAATCTTACTGTCTTTTAAACAAGACCA 2887
Qy 2869 tatctacaaaagacaggaatttgattaaacaaatataataacactaaatagaagaatgcttc 2928
Db 2288 TATCCACAAAAGACCAGGATTTGATTAACAGTATTAATACATTAATAGAGAANAATGCTTC 2347
Qy 2929 aggtcctcatctatattgtgggtgagcgttattgtacctggagtgggaaaatgtgaccaaaag 2988
Db 2348 AGGTCTCATCTATATTGTGGGAGAACGTTATGTACTTGGAGTGGGAAATGTTTACCAGAG 2407
Qy 2989 aagaggtcacaaatgagagaaaatcattcacttgccttgcattgaaccatgcacacagtg 3048
Db 2408 AGGAGGTTATAATAGAGAGATTTACTCACTTACTTGTTCATTTGAGCCCATGCCACAGTG 2467
Qy 3049 ccattgccaataatttaccctgagaatgaaataatgaaactggttagagaatgtcataa 3108

YNHLYPLHLITMAHMLQIITLTDLSPCPPLAEGEEDSEEARCASAFEVESQHTDG
LITCCGAPWHYWLNSLRNGITPYLRCAALLFHYLLGVAPPPBELFANSAGEFALSCLYL
SLPTNLFLEFWEYWDITIRPLLQRMWGDPAALLSKLKKSAVRYPRKRNLSLIELPEDIYS
CLLNQASHFRPSRADDKRPVLCFLGAILCSQNICQEIIVNGEVEGACVYFHALHC
GAGVCFILKIRECRVVLVEGKARGCAYPAPYLDEYGETDPGLKRGNPLHLRSRERYKLL
HLVWOHCIIIEIARSQETNQLMFENQOLL"
5389..6395
/gene="ubr1"
polyA_site
5429..5434
/gene="ubr1"

3'UTR
BASE COUNT 1802 a 1376 c 1519 g 1698 t
ORIGIN

Query Match 70.3%; Score 4435.4; DB 10; Length 6395;
Best Local Similarity 86.3%; Pred. No. 0;
Matches 4997; Conservative 0; Mismatches 721; Indels 69; Gaps 6;

QY 590 ttccagggccgfcgtaaaagtgtcctcctgtctc-tccgacgggcccacaggtttccgct 648
|||||
Db 8 TTCAGGGCCGTCGTAAAAAGTGTGFCCTGTGCGCTGCGGGCGGCCACAGGTTCCGCT 67
|||||
QY 649 tgccctctggcgggggtcgccactgcagcggtcagtttccctcaagatggcgagcagg 708
|||||
Db 68 AGCTGCGCGCGGGGTGCGGAACTGCGGGCGTTCGTTTCCCTTAAGATGGCGGAGAG 127
|||||
QY 709 aggcctggaggtactgagagatggaataatcagcgcgagtgatcccccagacccctcagctc 768
|||||
Db 128 AGATGACGCGCGCGGAGAGATGGAGCTCAGCCCGGAGGCTCCCTTGGCCCGCAGCGGC 187
|||||
QY 769 tggcatcttggtgggatacgaagtgtattttatctactcttcttcttgcattcttggcac 828
|||||
Db 188 CGGCATCGGTGGGTATGAGTACGAAAGTGTATTTATATCTGTTTCTTACATCATTTGGCAC 247
|||||
QY 829 aattggtgccgaaatttacttctgtgaaatggaccagacttggaaaagcagagagaaa 888
|||||
Db 248 AATTAGTCCCAAAATTTATTTGTGAGATGGACCCAGATTTGGAAAGCAGAGAGA 307
|||||
QY 889 gtgtcaaatgtcaatattctactccactggaaatggtacttatttggagaagatccagata 948
|||||
Db 308 GTGTACAGATGTCAATACTACTCTCTTTGGAGTGGTACTTATTTGGAGAGGATCCGGATA 367
|||||
QY 949 ttgtcttagagaaattgaagacagatggagcatttcagcttttggaggaggttttcaaaa 1008
|||||
Db 368 TTTGCTTAGAGAAATTAATAACACAGTGGAGCGTTCAGTGTGTGGGAAGGTTTTCAAA 427
|||||
QY 1009 gtggagagacaaccttattcttcaggaggtgtgcaattgatccaatgtgtactctgta 1068
|||||
Db 428 GTGGAGAAACAACATATTCCTGTAGGGATTGTCAATTTGATCCAACGTGTGTCTCTGTA 487
|||||
QY 1069 tggactgtcttcaggacagatgttcaataaaatcatcgtttacaagatgcatacttctactg 1128
|||||
Db 488 TGGACTGTCTTCAAAAGTAGTGTTCATAAAACCATCGTTTACAAGATGCATATCTTACTG 547
|||||
QY 1129 gaggagggtctctgactgtgagacacagaggaatggaaaactggcccttttggtaaa 1188
|||||
Db 548 GAGGGGGCTTCTGTACTGTGGAGACACAGAGCGGTGGAAAACTGGCCCTTTTGTGTGG 607
|||||
QY 1189 atcatgaacctggaagacagactactataaaagagaattcacgctgtccgttgaatgaag 1248
|||||
Db 608 ATCAGAGCGCTGGAGAGCAGGATCTACAAAAGAGAGCTTACATTTGCCCATTTGAATGAAG 667
|||||
QY 1249 aggtaatgttccaagccaggaataatttcttctcagtgatgaataatgtcgtagaatga 1308
|||||
Db 668 AGGTGATGCTCAAGCCAGGAGAAATATTCCTTTCGTTGATATAAATACATTGTAGAAATGA 727
|||||
QY 1309 ctatatgggaagagagaaagaaactgcctcctcgaactccagataaaggagagaaaaatgaaa 1368
|||||
Db 728 CTATATGGGAAGAAAGAAAGAAATTCCTTCTCAACTGTCAGATAGGGAGAAAAATGAAC 787
|||||
QY 1369 gatactattgtctcttcttcaatgatgaacacattcatatgaccacgtcatatacagcc 1428
|||||
Db 788 GATACATATTGTCTCTTTTCAACAGATGAGACCATTTTCGTATGATCATCATCATACAGTC 847
|||||

QY 1429 tacaagaagctcttgaactgtgagctcgagaggccaggtlgtcataccactgcatatgaca 1488
|||||
Db 848 TGCAGAGAGCTCTAGATTGCGAGCTTGCGAGGCGACAGCTGCACACGACTGCGCATCGACA 907
|||||
QY 1489 aagagggtcgctcgggcgtttaaagcggagccttatgtctgcttgcagagaacaaaggaag 1548
|||||
Db 908 AAGAGGTGCGCGGGCTGTCAAAGCAGGTGTGTATGCCACTTGCAGGAAGCAAGAGGAG 967
|||||
QY 1549 atataaagagtcattcagaaaaatgtctctcaacatccacttcattcatgaagtattacact 1608
|||||
Db 968 ATATAAAGAGTCACTCAGAGAACGCTCTCAGCACCCCTCCATGTGGAAGTGTCTCCACT 1027
|||||
QY 1609 cagagatatgtgctcatcagaataattgtcttgccttcttgggttctctggatgaacaaatta 1668
|||||
Db 1028 CCGTGGTATGGCTCACAGAAATTCGCTCTCGCGCTTGGCTTGGTGGATGAACAAAATTA 1087
|||||
QY 1669 tgaactattcaagtgacttttaggcagatcttttgcgaagcagccttagagaagaacctg 1728
|||||
Db 1088 TGAGCTATTCAAGTGACTTTTAGACAGATATTTTGGCAGGCTGCTCGTAGAAGAACCTTG 1147
|||||
QY 1729 actcggagaaatccctgtctcataaagcaggttaatgtcttgggatgcaagctttataaag 1788
|||||
Db 1148 GCTCTGAATCCCTGCTTATAAGCAGACTTAATGCTTTGGGATGCAAAACCTTTATAAG 1207
|||||
QY 1789 gtgcccgaagatccttcatgaattgtcttccagcagtttttttatggagatggaataaca 1848
|||||
Db 1208 GTGCCCTTAAGATCCTTCATGAATTGATCTTTAGTAGTATTTTATGTAGATGGAATACA 1267
|||||
QY 1849 aaaaactcttgcctatgaaatttgaagtattataaaactcagaaaaataataca 1908
|||||
Db 1268 AAAAACHCTTTGCTATGGAATTTGTGAAGTATATATAAACACTGCAGAAAGATACATCA 1327
|||||
QY 1909 gtgatgatcatgacagagatctctataactgcactttcagttcagatgtttactgttc 1968
|||||
Db 1328 GCGAGCACCAGGAGAGAGCATCTCCATAACCGCCCTGCTCGCGAGATGCTCACCGTCC 1387
|||||
QY 1969 ctactcgtgctcgacatcttattgaagagcagaattgttatctctgtctactgtaaacctc 2028
|||||
Db 1388 CGACTTTGGCCGGCATCTTATTGAAGAGCAGATGTTATTTCTGTCTATTACTGAAAGCC 1447
|||||
QY 2029 tgcctagaagtttacctgagtaacttgacagagaacaataaaattcaacttccagggttata 2088
|||||
Db 1448 TGCTAGAAGTTTACCTGAATACTTGGACAGGAACAATAAATTTCAACTTCCAGGGTTATA 1507
|||||
QY 2089 gcagagcaaatgtggagagatataatgcagttaattgtacctaagatatatccctgata 2148
|||||
Db 1508 GCCAGGACAAACTCGGGAAGAGCTTACCGAGTTATATGTGACCTAAAGTATATCTTGATTA 1567
|||||
QY 2149 gcaaaccccaaatatggcagagaagattaaagaatgcagttccttggaaaggttttctgactct 2208
|||||
Db 1568 GCAAGCCTGTCAATGGACAGAACGATTAAAGAGCGCAGTTTCCGCGGTCTT 1627
|||||
QY 2209 ttttgaagattcttaccctgtatgcaggaatggagaataatccgaagacaggttgggcaac 2268
|||||
Db 1628 TTCTGAAGATTCTTACCTGTATGCGGGAATGGAAGAAATCAGAGACAAGTTGGACAAC 1687
|||||
QY 2269 acattgaagtggatccctgattggaggctgccattgtctatcacagatgcaattgagaata 2328
|||||
Db 1688 ACATTGAAGTGGACCCCTGACTGGGAGGCTGCCATCGCTATACAGATGCAACTAAAGATA 1747
|||||
QY 2329 ttttactcatgttccaagagtgggtgtgtgtgtgataagaactcttacttgggttata 2388
|||||
Db 1748 TTTTGTGCTCATGTTCCAAAGAGTGGTGTGCTTGTGTATGAAGATCTCTTACTTGGTGGCTTATA 1807
|||||
QY 2389 aagaatgtcacaagctgtgatgagggtgcagttaccagtttccatctctagttaggacaagacag 2448
|||||
Db 1808 AAGAATGTCACAAAGCTGTAAATGAGGTGCAGTACAAATTTTCATGTCTCCATCCACAGACAG 1867
|||||
QY 2449 tagtacaactcgttgggacatagtttggaaaacaaagtcctacagatattctgaagatcttg 2508
|||||
Db 1868 TAGTGCAAATTGTGCGGTCTAGTCTGGAACCAAACTCTTACAAAAGTGTCTGAGGACCTTG 1927
|||||

Db 4088 CTTGGAGCACGTGCTGCTTCAACCATCCAGGCAATCGAAACACCTGTTGGGAGATGAAGGAA 4147
QY 4717 aacctctgttggagacacttcaaaaataggcagcatabaatggtctgaagaacttaataatgcagt 4776
Db 4148 AACCTTATTTGGAGCACTTCAAAATAGACAGCATACGCGTCTGAAGCGCTAATGCAGT 4207
QY 4777 ttgcagttgcacagagattacctgtccclcaggtcctgatacagaacaacatctggttcgtc 4836
Db 4208 TTGCAGTTGCACAGAGGCTTACCTGCGCCCTCAGGTCTCTGATACACAACATCTGGCTCGGC 4267
QY 4837 ttctatcagttgttcttctcaataaaaaatcagaagatacacacatgcctctctctctatag 4896
Db 4268 TCCTGTCAGTATCTCTCAACCTGCAATCACAANAATACACAGGCCCTCTGCTCTGG 4327
QY 4897 atctgtttctaatgttttggggggtgctgtgttagcatcccatctctgtattggagatgcc 4956
Db 4328 ATCTCTTCCATGTTCTGGTGGCGGCGAGTCTTAGCGTTCCCATCTCTGTATTTGGGATGACA 4387
QY 4957 ctgttgatctgaagcctctctcagtttagttcttctctatacaacacattatctcttccatt 5016
Db 4388 CCGTGGATCTGACGCGCTGCGCCACTTAGTTCTTCATATAACCCACCTCTATCTCTTCCATC 4447
QY 5017 tgatcacattggcacacatgcttcagatactactactacagtagaca-----cag 5064
Db 4448 TGATCACCATGGCGCATGCTTCAGATACTCTCTACACAGATACAGATCTGCTCCAG 4507
QY 5065 gactaccccttgctcaggttccaagaagacagtgaaagagctcattccgcatctcttcttct 5124
Db 4508 GGCGCGGCTTGTGAGGGTGAAGAGGATAGTGAGGAGGCTCGCTGTGCATCTGCTTTCT 4567
QY 5125 ttgcagaattctcaataatacaagtggtccatggtgtgatatctcctgctggtatt 5184
Db 4568 TTGTGAAGTGTGCGACACAGACGGCTCACTGGGTGCGGTGCTCGCGGTGTGTACC 4627
QY 5185 tgtgggtctcaactgaagaatggcatcaccccttactctgctgtgctgaatttttctcc 5244
Db 4628 TGTGGCTCTCCCTGAGAACGGCATCACCCCTTACCTCCGCTGTGCTGCACTCTTTTCC 4687
QY 5245 actattacttgggtaactcccgctgaggaactgcataccaattctgcagaagagagat 5304
Db 4688 ACTATTTACTTTGGAGTAGTCCGCCCTGAAGAACTGTTTGCCAATTTCTGCTGAAGGAGAT 4747
QY 5305 caagtcaactctgtagctatctatttaccatacaaatgttctcctgctcttccaggaaat 5364
Db 4748 TCAGTGCACCTCTAGCTATCTATCTTTTACCACCAAAATTTGTTCTCTGCTTTTCCAGGAAT 4807
QY 5365 attggataacttgaagccctgtctccagagtggtggtgcagatcctgccttactaaact 5424
Db 4808 ATTGGGATACCATTAAGCCCTTACTACAGAGGTGGTGGAGATCCTGCCCTTACTCAAGT 4867
QY 5425 gtttgaagcaaaaaacacagtggtcaggtacccttagaagaaagaaatagtttgatagagc 5484
Db 4868 CTTTGAAGCAGAAAGTGTGTGTGTCAGGTACCCCTAGAAAGAAATAGTTTGCATAGAGC 4927
QY 5485 ttctctgactatagctgctcctctgaatacaagcttctcatttcagggtgccacggtctg 5544
Db 4928 TTCTCTGAGGACTACAGCTGCTTCTTAATCAGGCTTCTCACCTTAGGTGTGCCAGGCTGTG 4987
QY 5545 cagatataagcaagaagcatctgtcctcctgcctttctgtgtgggctatactatgttctc 5604
Db 4988 CAGATATGAGCGAAGAGATCCCTGCTCTGCTCTTTTCTGTGGGGCATCTCTGTGTCTC 5047
QY 5605 agaacatttgcctccaggaaattgtgaacgagggaaggtgtgagcttgcatttttcaag 5664
Db 5048 AGACATCTGTTGCCAAGAAATAGTGAATGGGGAAGAGGTTGGAGCGTGGCTTTTTCATG 5107
QY 5665 cacttcaacttggagcggagctgctgatttctcttaaaaaatcagaagaatgccaggtggtcc 5724
Db 5108 CGCTTCAATTGCTGTGGAGTCTGCATTTTCTCTAAATCCAGAAATGCAGGCTGGTCC 5167
QY 5725 tgggtgaagtaaaagccagaggtgtgctctatccagctccttacttggatgaatatggag 5784
Db 5168 TGGTGAAGGAAAAAGCCAGAGGCTGTGCCTACCCAGCCCTTACTTTGGATGAATATGGAG 5227

QY 5785 aaacagacccctggcctggaagaggggcaaccccttcatttattctcgtgagcggtatcgga 5844
Db 5228 AAACAGACCCAGGGCTAAAGAGAGAAACCCACTTCTTATCTCGGAGCGGTATCGGA 5287
QY 5845 agctccatttggctctggcaacaacactgcattatagaagagattgctagagccaaagaga 5904
Db 5288 AGCTGCATTTGGTCTGGCAACAGCACTGCATTTATAGAAGAGATTGCTCGAGGCCAGAGA 5347
QY 5905 ctaatcagatgttatttggattcaactggcagcttactgtgagctcccaactctgcctccaag 5964
Db 5348 CTAAATCAGATGCTATTGATTTAACTGSCAGTTACTCTGAGCTTCAAGTTCTGCTCAAG 5407
QY 5965 acaatcacaatgacgacagtagtaaaagctgattcaaaatattggaacacttctctgag 6024
Db 5408 ACAATCATGAGTCACATCAATAATAAAGACTGATCTAAATTTCTAGAGAACTTTCTCAGG 5467
QY 6025 gctgggaaagtattggaggtcttttgcctcaggtccaggtccaggttccactacataaaaaat 6084
Db 5468 ACGGGGAAGTATTGGAGGGTCTTTTGATCCATGTCCAGATTCCACACACATTAATAAAAT 5527
QY 6085 atttctaatggagtattgcttcaattagcaaacatatgcttcacagagaaaaa-aggac 6143
Db 5528 ATTCTTAATGGAATATTGCTTCAATTATCAACATAGCTTCAAGSGAAACACAGAC 5587
QY 6144 atagatcaaatctgt-----ttt 6160
Db 5588 ATAGATTAATGTTTATGTTCTAGAACACATAAAGAAATGCTTGTTCATCCAAAGTGTCTAT 5647
QY 6161 atgtctagtagtattccaggaaatttattcccttcataaattgtctcattctattttttt 6220
Db 5648 TTCTGCTAATATTTCAGAAAACCTCTTTCCCTTCATAACTGCTGTAGTTCAATTTTCATAT 5707
QY 6221 catccacttgtagatgaagtcacgtcaaacagttgtgacacattttatgtgtggttaac 6280
Db 5708 CACCACCTGGTAAATGAGGTCACATTAAGCATTTAAGCATTTTGGACATTTCTCCATCTGGCTAAC 5767
QY 6281 tcttctgcaattttgtatttgggtttt 6307
Db 5768 ATCTCTGCACCTTGTATTGTTGTTT 5794

RESULT 4
AK027803
LOCUS
DEFINITION Homo sapiens CDNA FLJ14897 fis, clone PLACE1004743, weakly similar to PROBABLE N-END-RECOGNIZING PROTEIN.
ACCESSION AK027803
VERSION AK027803.1 GI:14042751
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens placenta cDNA to mRNA, clone_lib:PLACE1
clone:PLACE1004743.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (sites)
AUTHORS Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Aotsuka,S., Yoshikawa,Y., Matsumawa,H., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Negahari,K., Masuho,Y. and Sasaki,N.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2550)
AUTHORS Isogai,T. and Otsuki,T.
TITLE Direct Submission
JOURNAL Submitted (10-MAY-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)
COMMENT NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction; 5'- & 3'-end one pass sequencing and clone selection; Helix Research Institute (supported by Japan Key Technology Center etc.)

and Department of Virology, Institute of Medical Science,
University of Tokyo.
FEATURES
Location/Qualifiers
source
1. .2550
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="PLACE1004743"
/tissue_type="placenta"
/clone_lib="PLACE1"
/notes="cloning vector: pME18SFL3"
117..>2550
CDS
/note="unnamed protein product"
/codon_start=1
/protein_id="BAB55380.1"
/db_xref="GI:14042752"
/translation="MTIWEKEELPELOIREKNERYCYVLFNDEHHSYDHVYISLOR
ALDECELAQAQLHTTAIDKEGRRAVKAGAYAACQAEKEDIKSHSENVSOEPLHVEVLHS
EIMAHQKFLALRGSWNKIMSYSDFRQIFCQACLRDEEPSNPCLISRLMLWDALKY
KGAKTLHELIFSSPFMEYKKLFAMEFYKKQLOKEYISDDHDSISIRIVALSVM
FTVPTLARHLIEFONVISVITETLLEVLPEYLDNRNKNFQOYSQDKLGRIVAVICDL
KYLISKPTIWTERLMOFLEGRFSFLKILTCMOGMEIEIRROVGOHLEVPDPEAATA
IQMLKNILLMQEACDELELLVAYKECHKAVMRCSTSFISSTKITVWSGCHSLFT
KSTRSEDLVSLHPLSLAGLHVRSLRGVSRHFEVFSFEDQVEVLVEYPLRCL
VLVAQVAAEMRRNGLSLISQVEYYQDVCKREEMDKDIIIMLQIGASLMDPNKFFLLVY
LQRYELAEAFNKTISTKQDOLIKOYNLIIEMLQVLIYIVGERYVPGVGNVKEEVTM
RETILHLCTEPMPHSAIAKNLPENNETGLENVINKVATFKKPGVSGHGVYELKDES
LKDNMYFVHYSKTOHSAHEMOKKRRKOENKDEALPPPPPECFAPFSKVINLLNCD
IMMYILRTVEERAIIDSLNWTGEMLOMAFHILALGLEKQQLKAPESBEVTFDFTH
KASRLGSSAMNOMLLEKLGIPQLEGQDKDMITWIIQMFDTVRLRKRKSLIVATTSG
SESIKNDIEITHDK"
BASE COUNT 830 a 454 c 558 g 708 t
ORIGIN

Query Match 40.4%; Score 2546.8; DB 9; Length 2550;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2546; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1189 atcatgaacctggaagcagggtactataaagagaattcaacgtgtccgttgatgaag 1248
Db 1 ATCATGAACCTGGGAAGCAGGACTATAAAGAGAATTACGCTGTCGCTGGAATGAAG 60
QY 1249 aggtlaattgtccagcagaaataatttccttcagtgataaaatatgtcgtagaaatga 1308
Db 61 AGGTAATTGTCCAGCCAGCAGAAATAATTCTTCAGTGATAAATAATGCCGTAGAAATGA 120
QY 1309 ctaatgggaaggaagaaagactgcctcctgaactccagataagggagagaaatgaaa 1368
Db 121 CTATATGGGAAGAGAAAAAGAACTGCCTCTGAACCTCCAGATAAGGCAGAAAAATGAAA 180
QY 1369 gatactattgtctcttttcaatgatgaacacacattcatatgaccacgtcatatagacc 1428
Db 181 GATACTATTGTGCTCTTTCAATGATGAACACCATTCATATGACCACGTCTATATACAGCC 240
QY 1429 tacaaagactcttgactgtgactgcagagccaggttgcataccactgcccattgaaca 1488
Db 241 TACAAAAGAGCTCTTGACTGTGAGCTCGCAGAGCCCAAGTGTGCATACCACCTGCCATTGACA 300
QY 1489 aagaggtctcggggtgtttaaagcggagccttatgctgtctgcaggaagaaagaaag 1548
Db 301 AAGAGGCTCGGGCTGTTAAAGCGGAGCTTATGCTGCTTGGCTTGCAGGAAGGAAG 360
QY 1549 atataagagatcattcagaaaaatgtctctcaacatccactcatgtagaagtattacact 1608
Db 361 ATATPAAAGAGCTCATTCAGAAAAATGTCTCTCAACATCCACTTCATGTAGAAGTATTACACT 420
QY 1609 cagagattatggctcatcagaaatttgccttgcctgttgccttgcctgtgatacaaaaaatta 1668
Db 421 CAGAGATTATGGCTCATCAGAAATTTGCTTTGGCTCTTTGGTTCTTGGTATGAACAAAAATTA 480
QY 1669 tgagctattcaagtgcactttaggcagatcttttggcagcatgccttagagaagaaacctg 1728
Db 481 TGACCTATTCAAGTACTTTAGGCAGATCTTTTGGCCAGGATGCTTTAGAGAAAGAACTGT 540

QY 1729 actcggagaaatccctgtctcataaagcaggttaatgctttgggagtgcaaaactttataaag 1788
Db 541 ACTCGGAGAATCCCTGTCTCATAAAGCAGGTTAATGCTTTGGGATGCAAAAGCTTTATAAG 600
QY 1789 gtyccgtaagatcccttcataatgaattcagcaggtttttttatggagatggaataca 1848
Db 601 GTGCCCGTAAGATCCCTTCATGAATTCATCTTCAGCAGTGTTCATGGAGATGGAATACA 660
QY 1849 aaaaactctttctatggaattgtgaagtattataaaacaactgcagaaagaatatatca 1908
Db 661 AAAAACTCTTTGCTATGGAATTTGTGAAGTATTATAAACAACATGCAAAAGAAATATATCA 720
QY 1909 gtgatgatcatgacagaagtatctctataactgcacttcagttcagatgtttactgttc 1968
Db 721 GTGATGATCATGACAGAAAGTATCTCTATACAGCACTTTCAGTTCAGATGTTTACTGTTC 780
QY 1969 ctactctggctgcacatcttattgaagagcagaatgtatctctgtcattactgaaactc 2028
Db 781 CTACTCTGCTCGACATCTTATTGAAGACAGAAATGTTATCTCTGTCTCATTACTGAAACTC 840
QY 2029 tctcagaagtcttaccctgagttacttggacagaaacaataaaattccactctccaggttata 2088
Db 841 TGCTAGAAGTTTACCTGAGTACTTGGACAGGAACAATAAAATTCACCTCCAGGGTTATA 900
QY 2089 gccagacaattgggaagcagatatatgcagtaatatgtacactaaagatatctcctaatca 2148
Db 901 GCAGACAAATTTGGGAAGAGTATATGCAGTAAATATGTGACCTAAAGTATATCTCTGATCA 960
QY 2149 gcaaaccccaataatgacagaaagattaagaatgcagttccttgaaggttttctcgtact 2208
Db 961 GCAACCCACAAATATGACAGAAAGATTAAAGATGCAGTTCCTTTGAAGGTTTTTCGATCTT 1020
QY 2209 ttttgaagattcttacctgtatgcagggaattggaagaaatccgaagacacaggttgggcaac 2268
Db 1021 TTTTGAAGATTCTTACCTGTATGTCAGGGAATGGAAGAAATCCGAAGACACAGGTTGGCAAC 1080
QY 2269 acatgaagtgcactgcattggaggctgccattctctacagatgcattatgaagata 2328
Db 1081 ACATTGAAGTGTGATCCTGATGGGAGGCTGCCATTGCTATACAGATGCAATTGAAGAAATA 1140
QY 2329 ttttactcatgttccaagagtggtgtgcttgcagtaagaaactcttacttgcgtgcttata 2388
Db 1141 TTTTACTCATGTTCCAAAGAGTGGTGTGCTGTGATGAAGAACTCTTACTTGTGCTTATA 1200
QY 2389 aagaatgtcaaaaagctgtgtagaggtgcagttaccagtttcatatctatgtagcaagacag 2448
Db 1201 AAGAATGTCACAAAGCTGTGATGAGGTGCAGTACCAGTTTCATATCTAGTAGCAAGACAG 1260
QY 2449 tagtacaatcgtgtgacatagttttgaacaaagtcctacagatctcagagatcttg 2508
Db 1261 TAGTACAATCGTGTGACATAGTTTGGAAACAAGTCCACAGAGTATCTGAGAGATCTTG 1320
QY 2509 taagcatatcatctgcactctctaggaacctgtgctgtcttcacgtacgtttaaagcagc 2568
Db 1321 TAAGCATACATCTGCCACTCTCTAGGACCCCTTCTGCTGTTCATGTACGTTTAAGCAGGC 1380
QY 2569 tgggtgctgtttcaagactgcagtaattgtgtctttttgaggactttcaagttagaggtac 2628
Db 1381 TGGGTGCTCTTTCAAGACTGCATGAATTTGTGTCTTTTGGAGGACTTTTCAAGTAGAGGTAC 1440
QY 2629 tagtgaataatcctttacagttgtctggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2688
Db 1441 TAGTGGAAATATCTTTTACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1500
QY 2689 gaagaaatgagctgtctcttattagccaggtgttttattaccagaagtgttaagtgcagag 2748
Db 1501 GAAGAAATGGACTGTCTCTTATTAGCCAGGTGTTTATTACCAGAGATGTTTAAGTGCAGAG 1560
QY 2749 aagaaatgatataaagatacatcatcttcagattgtgtgcatctttaaattgagatccca 2808
Db 1561 AAGAAATGTATGAATAAGATATCATCTGCTCAGATGCTGTGATGCTTCTTAAATGGATCCCA 1620

QY 2809 ataaagtcttctgtacttggtacttcagaggtatgaacttgcgaggtcttttaacaaagacca 2868
|||||
Db 1621 ATAAAGTCTTGTTACTTGTTACTTCAGAGGTATGAACCTTGCCGAGGCTTTTAAACAAGACCA 1680
QY 2869 tatctacaaaagaccaggtattgattaaacaataataacactaaatagaagaatgcttc 2928
|||||
Db 1681 TATCTACAAAAGACCAGGATTTGATTAAACAATATAACACTAATAGAGAATGCTTC 1740
QY 2929 aggtctctatctatatgttggtgagcgttatgtacactggagtggaataatgtgacaaaag 2988
|||||
Db 1741 AGGTCCTCATCTATATGTGGTGGAGCGTATGTACGTGGAGTGGGAAATGTGACCAAAAG 1800
QY 2989 agaggtcacaaatgagagaaatcattcacttgccttcttgactgaacccatgccacagtg 3048
|||||
Db 1801 AGAGTCAATGATGAGAGAAATCACTTCTGCTTTGTCATTTGAACCCATGCCACACAGTG 1860
QY 3049 ccattgccaataattaccttgagaatgaaataatgaactggcttagagaatgtcataa 3108
|||||
Db 1861 CCATTGCCAAAATTTACCTGAGAATGAAATTAATGAACCTGGCTTAGAGAATGTCTATA 1920
QY 3109 acaagtgccacatttaagaaacccgggtgtatcaggcccatggagtttatgaactaaaag 3168
|||||
Db 1921 ACAAAAGTGCCACATTTAGAAACCCAGGTGTATCAGGCCATGGAGTTTATGAACATAAAG 1980
QY 3169 atgaatcactgaagaacttcaatatatacttcttattcattactccaaaacccagcatagca 3228
|||||
Db 1981 ATGAATCACTGAAGACTTCTCAATATATGCTTTTATCATCTCCAAAACCCAGCATAGCA 2040
QY 3229 aggtgaaacatatgcagaagaagagagaaacaagaacaaagatgaagcattggcgc 3288
|||||
Db 2041 AGGCTGAACATATGCAGAGAAAGAGAGAAACAGAAACAAAGATGAAGCATTCGCCG 2100
QY 3289 caccaccactctctgaattctgcccgtgttccagcaagtgatgaactctcaactgtg 3348
|||||
Db 2101 CACCACCACTCTGAATTTCTGCCCTGCTTTCAGCAAGTGATTAACCTTCTCAACTGTG 2160
QY 3349 atatcatgatfacattctcagaccgtatttgagcgggcaatagacacagattctaaact 3408
|||||
Db 2161 ATATCATGATGTACATTTCTCAGGACCGTATTTGAGCGGGCAATAGACACAGATTTCTAACT 2220
QY 3409 tgtgacqaagggatgctccaaatggcttttcattatcttctggttgggtttactagaag 3468
|||||
Db 2221 TGTGACCGAAGGATGCTCCAAATGGCTTTTCATATTTCTGGCATTTGGGTTTACTAGAAG 2280
QY 3469 agaagcaacagcttcaaaaagctcctgaagaagaagtaacatttgacttttatcataaag 3528
|||||
Db 2281 AGAAGCAACAGCTTCAAAAAGCTCCTGAAGAAGAAGTAACATTTGACTTTTATCATAAGG 2340
QY 3529 ctccaagattgggaagttcagccatgaatatatacaaatgcttttggaaaactcaaaagaa 3588
|||||
Db 2341 CTTCAAGATTGGGAAGTTTCAGCCATGAATATACAAATGCTTTTGGAAAACCTCAAAAGAA 2400
QY 3589 ttccccagttagaagcccaagagcatgatgaactggtgacttcagatgtttgacacag 3648
|||||
Db 2401 TTCCCCAGTTTAAAGCCCAAGAGGACATGATTAACGTGGATCTTCAGATGTTTTCACACAG 2460
QY 3649 tgaagcgtatgaagaaaaattctgttttaattgttagcaaccacatcagatcggaatcta 3708
|||||
Db 2461 TGAAGCGGTTAAGAGAAAAATCTGTTTAAATGTAGCAACCCACATCAGGATCGGAATCTA 2520
QY 3709 ttaagaatgatgagattactcatgataaag 3738
|||||
Db 2521 TTAAGAATGATGAGATTACTCATGATAAAG 2550

RESULT 5
LOCUS AR030785 1001 bp DNA PAT 29-SEP-1999
DEFINITION Sequence 2 from patent US 5861312.
ACCESSION AR030785
VERSION AR030785.1 GI:5943999
KEYWORDS
SOURCE Unknown.

ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1001)
AUTHORS Varshevsky, A. and Kwon, Y. Tae.
TITLE Nucleic acid encoding mammalian UBR1
JOURNAL Patent: US 5861312-A 2 19-JAN-1999;
FEATURES Location/Qualifiers
source
1..1001
BASE COUNT 363 a 186 c 205 g 247 t
ORIGIN

Query Match 15.8%; Score 996.2; DB 6; Length 1001;
Best Local Similarity 99.7%; Pred. No. 2.6e-240;
Matches 998; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2799 atgatcccaataaagttcttcttacttgacttcagaggtatgaacttccgaggtttt 2858
Db 1 ATGGATCCCAACAAGTTCTTGTACTGTACTTCAGAGGTATGAACCTGCCGAGGCTTTT 60
QY 2859 aacaagaccatctacaaaagaccaggtttgattaaacaataataacactaaatagaa 2918
Db 61 AACAAAGACCATATCTACAAAAGACCAGGATTTGATTAAACAATATAACACTAATAGAA 120
QY 2919 gaattgtctcagggtcctcatctatatattgtggtgagcgttatgtacctggagtggaat 2978
Db 121 GAAATGCTTTCAGGTCCTCATCTATATTGTGGTGGAGCGTTATGTACCTGGAGTGGGAAAT 180
QY 2979 gtgaccaagaagaggtcacaatgagaaatcattcacttcttctgattgaacccatg 3038
Db 181 GTGACCAGAAGAGGTACAAATGAGAGAAATCATTCACCTTGTTCATTTGAACCATAG 240
QY 3039 ccacacagtgccattgccaaaatttacctgagaatgaaataatgaacactggtcttagag 3098
Db 241 CCACACAGTGCATGCCAAAATTTACTGAGATGAAAATTAATGAACCTGGCTTAGAG 300
QY 3099 aatgtcataaacaagtggtccacatttaagaaacccaggtgtatcaggccatggagttat 3158
Db 301 AATGTCATAAACAAGTGGCCACATTTAAGAAACCCAGGTGTATCAGGCCATGAGTTTAT 360
QY 3159 gaactaaagatgaatcaactgaagaacttcaatatgtacttttatactactccaaaacc 3218
Db 361 GAATTAAGAAGATGAATCACTGAAGACTTCAATATGTACTTTTATCATTTACTCCAAAACC 420
QY 3219 cagcatgaagaggtgacacatatgcagaagaagaagaagaacaaagaacaaatgaa 3278
Db 421 CAGCATAGCAGGCTGAACATATGCAGAAGAAAAGAGAAAACAGAAAACAAAGATGAA 480
QY 3279 gcattgcgcgccaccacacctcctgaattctgacctgctttcagcaaatgatttaacctt 3338
Db 481 GCATTGCCGCCACCACCACCTCCTGAATTTCTGCCCTGCTTTTCAGCAAAAGTATTAACCTT 540
QY 3339 ctcaactgtgatatcatgatgtatacttcaggaccgtatttgagcgggcaatagacaca 3398
Db 541 CTCAACTGTGATATCATGATGTACATTCAGGACCGTATTTTGGCGGCAATAAACACA 600
QY 3399 gattcctaacttggagcgaaggtgctccaaatggcttttcatattctggcatctgggt 3458
Db 601 GATTCTAACTTGTGGACCGAAGGATGCTCCAAATGGCTTTTTCATATTTCTGCAATTTGGGT 660
QY 3459 ttactagaagaagaacacagcttcaaaaagctcctgaagaagaagaatgaacatttgacttt 3518
Db 661 TTACTAGAAGAGAAGCAACAGCTTCAAAAAGCTCCTGAAGAAGAAGTAACATTTGACTTTT 720
QY 3519 tatcataaggtctcaagattgggaagttcagccatgaatatatacaaatgcttttgaaaaa 3578
Db 721 TATCATAGGCTTCAAGATTGGGAAGTTTCAGCCATGAATATACAAATGCTTTTGGAAAAA 780
QY 3579 ctcaaaaggaattccccagtttagaagggcagaggaacatgaataacgtgatacttcagatg 3638
Db 781 CTAAAGGAATTTCCCACTTTAGAAGGGCAGGAAGGACATGATAAACGTGGTACTTTCAGATG 840


```

/protein_id="AAC39845.1"
/db_xref="GI:3170889"
/translation="MDPNFLLLVLQRYELAEAFNKTISTKDQDLIKQYNTLTLEMLQ
VLITYIGERYVGVGVNTEVEYTMREI IHLCTEPMPSAIAKNNPENNETTGLNV
INKVATFKPGVSGHGVKDESKDFNMYFYHSKTOHSAEHMOKKRRQKQENDE
ALPPPPPPFCFAFSKVINLNCIDIMYILRTVFERAINTDSNLWTEGMLQMAFHILA
GLLEEKQDQKAPEEVFTFYHKASRUGSSAMTOMLEKLGKLPQEGQKDMITW
ILOMFTVKRLRKEECVLIVATVSGESIKNDEITHDKAEKRKRAEAAHLRHKQIM"
363 a 185 c 204 g 247 t
BASE COUNT
ORIGIN

```

	Query Match	15.8%;	Score 994.2;	DB 9;	Length 999;
	Best Local Similarity	99.7%;	Pred. No. 8.2e-240;		
	Matches	996;	Conservative 0;	Mismatches 3;	Indels 0; Gaps 0;
Qy	2799	atggatcccaataagttcttgttactcgttactcagaaggtatgaacttgcgcgaggtcttt	2858		
Db	1	ATGATYCCCACAAAGTCTCTGTGTACTGTGTAATCTCAGAGGTATGAACCTGCGAGGCTTT	60		
Qy	2859	aacaagaccatattctaaaaagaccaggatttgattaaaaaataataacactaataagaa	2918		
Db	61	AACAAGACCATATCTACAAAGACCAGGATTTCATTAACAATATATACACTAATAGAA	120		
Qy	2919	gaatgcttcaggctcctcatctatatattgtgggtgagcgttatgtaccctggagtggaagt	2978		
Db	121	GAATGCTTCAGGTCCTCATCTATATTGTGGGTGAGCGTTATGTACCTGGAGTGGGAAT	180		
Qy	2979	gtgaccaaagaagaggttcacaatgagagaaatcattcacttctgtcttcatttgaacccatg	3038		
Db	181	GTGACCAAGAAGAGGTCAATGACAGAAATCATTCCTCTGCTTTGCAATGAACCCATG	240		
Qy	3039	ccacacagtgccattgcctcaaaaatttaccctggaatgaaaataatgaacctggcttagag	3098		
Db	241	CCACACAGTGCCATTGCCAAAAATTTACCTGAGANTGAAATATGAAACTGCTTAGAG	300		
Qy	3099	aatgtcataaacaagtggccacatttaagaacccaggtgtcatcagccatggagtttat	3158		
Db	301	AATGTCTATAAACAAAGTGGCCACATTTAAGAAACCAGGTGTATCAGSCCATGGAGTTAT	360		
Qy	3159	gaactaaagatgaatcactgaaagacttcaatatgtacttttatcattactcctcaaaacc	3218		
Db	361	GAACTAAAAGATGAATCACTGAAAGACTTCAATATGTACTTTTATCATTACTCCAAACC	420		
Qy	3219	cagcatagcaaggctgaacatatgtcagaagaaaagagaaaacaagaacaagaatgaa	3278		
Db	421	CAGCATAGCAGGCTGAACATATGCAGAGAAAAGGAGAAAACAGAAAACAAGATGAA	480		
Qy	3279	gaattgccgccaccacacacctcgtgaattctgcctgcttccagcaaatgtataacctt	3338		
Db	481	GCATTTGCCGCCACACACCTCCTGAAATCTTGCCCTGCTTTACGAAAAGTGATTAACCTT	540		
Qy	3339	ctcaacttgatatacatgatgtacattctcaggacccgtatttgagcggcaatagacaca	3398		
Db	541	CTCAACTGTGATATCATGATGATACATCTCAGGACCGTATTTGAGCGGGCAATTAACACA	600		
Qy	3399	gattctaaacttgggaccgaaggatgctccaaaatggcttttcatattctcgtatgggt	3458		
Db	601	GATTCCTAACTTTGTGAGCCAGAGGATGCTCCAAATGGCTTTTCATATTCGGCATTTGGGT	660		
Qy	3459	ttactagaagaagaacacagcttcaaaaagctcctgaagaagaagtaaacatttgacttt	3518		
Db	661	TTACTAGAAGAAAGCAACAGCTTCAAAAAGCTCCTGAAGAAGATACATTTTGACTTT	720		
Qy	3519	tatcataaggcttcaagatltgggaagttcagccaatgaaatatacaaatgtcttltggaaaaa	3578		
Db	721	TATCTAAGAGGCTTCAGATTTGGGAAGTTCAGGCCATGAATATACAAATGCTTTTGAAAAA	780		
Qy	3579	ctcaagaatctccagttagaagccagaaggacatgataacgtggatacttccagatg	3638		
Db	781	CTCAAGGAATTTCCCAGTTTGAAGGCCAGAGGACATGATACGTGGATCTTCAGATG	840		
Qy	3639	tttgacacagtgaagcqaatlaaagaagaaaatcttqtttaatttgaacacaccacatcaga	3698		

Db	841	 TTTGACAGTGAAGCGATTAAAGAAAAATCTTGTAAATGTAAGCAACCATCAGGA	900
Qy	3699	tcgaatctattaagaatgatgagattactcatgataaagaaaaacagaacgaaaaga	3758
Db	901	TCCGGAACTATTAGAATAATGATGAGATTACTCATGATAAAGAAAACGACAACGAAAAAGA	960
Qy	3759	aaagctgaagctgtaggtacatcgccagagaagatcatg	3797
Db	961	AAAGCTGAAGCTGTAGGCTTCATCGCCAGAGATCATG	999
RESULT	8		
LOCUS	AB002347	6158 bp mRNA PRI	13-FEB-1999
DEFINITION	Human mRNA for KIA0349 gene, partial cds.		
ACCESSION	AB002347		
VERSION	AB002347.1	GI:2224638	
KEYWORDS	KIAA0349.		
SOURCE	Homo sapiens male brain cDNA to mRNA, clone_lib:pBluescriptII SK plus clone:HGL561.		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 6158)		
AUTHORS	Nagase,T., Ishikawa,K., Seki,N., Nakajima,D., Ohira,M., Miyajima,N., Kotani,H., Nomura,N. and Ohara,O.		
TITLE	Direct Submission		
JOURNAL	Submitted (28-MAR-1997) to the DDBJ/EMBL/GenBank databases. Nobuo Nomura, Kazusa DNA Research Institute, Gene Structure 1; 1532-3 Yana, Kisarazu, Chiba 292, Japan (E-mail:cdnainfo@kazusa.or.jp, URL:http://www.kazusa.or.jp, Tel:+81-438-52-3930, Fax:+81-438-52-3931)		
REFERENCE	2 (sites)		
AUTHORS	Nagase,T., Ishikawa,K., Nakajima,D., Ohira,M., Seki,N., Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.		
TITLE	Prediction of the coding sequences of unidentified human genes. VII. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro		
JOURNAL	DNA Res. 4 (2), 141-150 (1997)		
MEDLINE	97349984		
FEATURES	Location/Qualifiers		
source	1..6158		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="HG1561"		
	/clone_lib="pBluescriptII SK plus"		
	/sex="male"		
	/tissue_type="brain"		
gene	1..3828		
	/gene="KIAA0349"		
CDS	<1..3828		
	/gene="KIAA0349"		
	/codon_start=1		
	/protein_id="BAA20806.1"		
	/db_xref="GI:2224639"		
	/translation="LKYLVLSIKPTWSDLELRKFLEGFDAFLLELLKCMQMGPITRQV GOHIEMEPNEAEFTLOMLKTHVTSMQDWCSDEKVLAYKKCLAVLMQCCHGQTVD GEOPITSLICGHVSFETRYCVSOEKVSIHLPRQLLAGLHVLLSKSEVAKYKPPELLP SELSPMDNHEPLRLVLCAQHAGWMRRNGFSLVNQIYYIHNVKCRMFMDKVDMVL QTGSMMDNHFLIMLSRFELYQIFSTDYDKRFSSEITHDVKQONNTLTTEMLYL ILMVGRSGPGVGOWNADEIKREIIHOLSFKPMASSELVKSLEPENKMGESVI EVAHFKPKGLTRGRGVNELKPKCAEFNFYLTFHSRAEOSKAEEAQRLKRONREDTA LPPPVLPFCPLAFSLANSIDVMLCIMGTLFWAVEHNGYAWSESMLQRVHLHGM ALOEKEQHLENYTEHVHTTFTQIKSGPEAKNPSPISLAMLETLQNAFYLEVHKDM IRWLKTFNAVKKMRESSPTSVAETEGTIMEESSRDOKDKARKRKAEIARLRREKIM AOMSEMORHFIDENKLFOTLELDASTSVLDHSPVASDMTLTALGPTQTQVPSQRO FYTCILCOEQEVKVESRAMVLAAFORSTVLSKNRSKFIOAPEKYDPLFMHPDLSCG THTSSCGILMAHCWORFDSVOAKEERQORRUHTSYDVENGELFCLPCLECLSNLY IPLLPPRIFFNNRLNSDPNLQWIRTISOQIKALOFLRKEEDTPNNASTNKSEN ILOPEGRPFPRFRPKIPQSISEKMLTFTGYTKVGLKHVPHEEDPRVPTMCWGSC AVTIYSIEILSDOEOKPLRGPLUPKDFSLTRFAAAHWTVASVVQGHCKLEA SLVPNDSHFEPLCILIDNMFIHVLGVLAEPALOCODESCISIGWDIHHIHPMAH		

IIIOILLTCTEENGMDENPPCEESAVLALYKTLHQYTGSALKKEIPSGWHLRSVRA
GIMPELKCSALFPYLVNGVSPDDIQVPGTSHEFLCSYLSLNPNNLJCLQFENSEIMN
SLIESCRNSEVRYLEGRSDAIRYPRESKNLILNPEDYSLLINQANFSQPKSGGDK
SRAPTCLVLGCSULCSQSYCCQTELEGEDGACTAHTYSCGSGVGIPLRVRECOVLFL
AGKTKGCFYSPVLYDDYGETDQGLRRGNPLHLCKERFKIKQLWHQHSVIEIGHAQE
ANQTLVGDWHL"

BASE COUNT 1832 a 1199 c 1320 g 1807 t
ORIGIN

Query Match 15.6%; Score 981; DB 9; Length 6158;
Best Local Similarity 55.9%; Pred. No. 2.1e-236;
Matches 2157; Conservative 0; Mismatches 1610; Indels 90; Gaps 12;

Qy	2130	ctaaagtatactctgatacagcaaacccacacaatattgacagagaagattlaagaatcgacttc	2189
Db	1	CTCAAGTATGTTAATAGCAAAACCAACTGAATGGTCAGATGAGCTGAGCGCAAGAGTTC	60
Qy	2190	cttgaaggttttgcattcttttgaagattcttacctgtgtatcgagggaatggaagaatc	2249
Db	61	CTAGAAGGGTTGATGCTTTTGGAAATTACTAAAAATGATGCAGGGAATGGATCCAATT	120
Qy	2250	cgaagacaggttgggcaacacattgaagtgatcctgattgaggaggctgccattgctata	2309
Db	121	ACAGTCAAGTAGGACACACATATTGAAATGGAAACAGAGTGGGAAGCAGCCCTTCACACTA	180
Qy	2310	cagatgaattgaagaataatttactcatgttccaagagtggtgtctgtgtagaagaa	2369
Db	181	CAATGAANATTACACATGTCATTTCAATGATGAGGACTGGTGTGCTTCAGATGAAAAA	240
Qy	2370	ctcttacttggcttataaagaatgcacaaaagctgtgatgaggtgcagtlaccagtttc	2429
Db	241	GTGTTAATCGAAGCTTACAGAANAATGCTCGCTCTACTGTATGCAGTGCATGGTGTAT	300
Qy	2430	atatctagtagcaagacagt----agtaacatcgtgtggaacatagttggaaacaaagtcc	2486
Db	301	ACTGATGGTGAACAGCCCAATCACACTAAGCATTTTGTGGACATTCAGTGGAAACATCAGA	360
Qy	2487	tacagatctctagagatctgtgaagatacatctgccactctctagacccttgcgtgt	2546
Db	361	TACTGTGTTCCCAAGAAAAGTTAGCATTCACCTCCAGTTTCTCGCTTACTTGCAGGT	420
Qy	2547	cttcatctagctttaagcaggtgggtgctgtgtttcaagactgcataagatttgtctttt	2606
Db	421	TTACATCTATTATTAGCAAAAAGGAAGTGCATATAAAATTTCCAGAGCTCTACCTCTA	480
Qy	2607	gaggacttcaagtagaggtagtggataatctcttcaotgtgtcgtgtgtgtgtgcc	2666
Db	481	AGTGAACCTTAGCCCCACCCATGTTGATAGAACACACCTCTTAGATGTCTTGTGTGTC	540
Qy	2667	caggttgtctgaagatgtgcaagaataaggactgtctcttattagccaagtgtttat	2726
Db	541	CNAGTACATCCGGAATGTGGGAAGAAATGGGTCTCTCTAGTAAACAGATTTATTAC	600
Qy	2727	taccaagattgaagtcagagaagaattgtatgataaagatacatcatcttcagatt	2786
Db	601	TACCAATATGTAATCCAGACGAGATGTTTGCACAAAGATGATAGTAATGCTTCAGACA	660
Qy	2787	ggtgcatttaatggtatcccaataaagtcttctgttactggtaacttcagagagtgaactt	2846
Db	661	GGTGTCCTCATGATGGATCCAAATCATTTCTCTGATGATCATGCTCAGCCGCTTTGAAC	720
Qy	2847	gcagagcttttaacaaagaccatatctacaaa-----gaccac	2885
Db	721	TATCAGATTTTCTAGTACTCCAGACTATGGAAAAAGATTTAGTTCTTGAGATTTACCCATAAG	780
Qy	2886	gatttgattaaacaataatacactaatagaagaataatgcttcaggctcctcatctatatt	2945
Db	781	GATCTGTTGTCAGACAGAAATFACTCTTAATAGAAGAAATGCTTATACCTATTATATGCTT	840
Qy	2946	gtgggtgagcgttatgtacctggagtgaggaaatgtgacaaagaagaggtgcacaatgaga	3005
Db	841	GTTGGAGAGAGATTTAGTCTCTGGAGTTGGACAGCTAAATGCTACAGATGAATCAAGCGA	900

Qy	3006	gaaatcattcacttgcttgcattgaaccatgccacacagtcgcatggccaaaattta	3065
Db	901	GAGATTATCCATCAGTTGAGTATCAAGCCTATGCTCATAGTAATGGTAAAGTCTTTA	960
Qy	3066	cctgagaatgaaaaataatgaaactggcttagaagaatgcataaacaagtgggcacattt	3125
Db	961	CCTGAAGATGAGAACAGGAGACTGGCATGGAGTGAATCGAAGCAGTTGCCATTTTC	1020
Qy	3126	agaaaccaggtgtatcaggccaatgagttttatgaactaaagaatgaatcactcaagaac	3185
Db	1021	AGAAAACTGGATTAACAGACGAGGCATGTATGAATGAACAGAGATGTGCCAAGAG	1080
Qy	3186	ttcaatatgtactttttatcattactcaaaacccagcatcaagagctggaacatatcag	3245
Db	1081	TTCAACTTGTATTCTTCTACTCTTTCAAGGGCAGAACAGTCCCAAGGAGAGACGCA	1140
Qy	3246	agaaaaaggagaaaaacaagaaagatgaagcattgcccaccaccacccctcctgaa	3305
Db	1141	CGSAAATTGAAAAGACAAAATAGAGAAGATACAGCACTCCACCTCCGGTGTTCCTCCA	1200
Qy	3306	tcttgccctgtcttcagcaaaagtattaaaccttctcaactgtgatcatcatgatgtacatt	3365
Db	1201	TTCTGCCCTCTGTGTTCCAAAGCCTGGTTAACATTTTGCAGTCAGATGTCATGTTGTGCATC	1260
Qy	3366	ctcaggacccgtattttagcgggccaatagacacagattctcaactgttgaccgcaagggatg	3425
Db	1261	ATGGGACAAATTTCTGCAATGGGCTGTGGAACATAAATGGATATGCCTGGTCAGAGTCCATG	1320
Qy	3426	ctccaaatggcttttcatattctggcattgggttttactagaagaagaacacagcttcaa	3485
Db	1321	CTGCAAAAGGCTGTACATTTAATTTGGCATGGCACTACAAGAAGAAAAACAACATTTAGAG	1380
Qy	3486	aaagctcctggaag--aagaagtaacatttgaacttttatcataaggcttcaagattggga	3542
Db	1381	AATGTCAACGGAAGAGCATGTAGTAACATTTACCTTCTCAGAAAGATATCAAAACCTGGT	1440
Qy	3543	agttcagccatga-----atacaaatgcttttggaaaaactcaagaagaattccc	3593
Db	1441	GAAGCCCAAAAATTTCTCTAGCATACTAGCTATGCTGGAAACACATACAAAATGCTGCC	1500
Qy	3594	cagttagaagggccagaaagacatgataacgtggatacttcagatgtttgacacagtgtaag	3653
Db	1501	TACCTAGAAAGTCCACAAAGACATGATTGCGTGGATATTGAAGACTTTTAAATGCTTTAAA	1560
Qy	3654	cgattaaagggaaaaactctgttttaattgttagcaacacatacaggatcgggaatctattaag	3713
Db	1561	AAAGATGAGGGAGA-----GTTCCCTTACCAGTCCCGTGGCAGAGACAGAAAGAACCATTA	1614
Qy	3714	aatgatggatLactcatgataaagaaaaagcagaaagcaaaagaaagctgaaagctgct	3773
Db	1615	ATGGAAGAGAGTTTCAAGGGACAAAGACAAAGCTGAGAGGAGAGAAAGACAGATTTGCC	1674
Qy	3774	aggctacatcgcgcagaagatcattggctcagatgtctgctcttacagaaaaacttcattgaa	3833
Db	1675	AGACTCGCAGAGAAAGATCATGGCTCAGATGCTGCTGAAATGCGACGCGCAITTTTATTGAT	1734
Qy	3834	actcataaactcatgtatgacaaatgcatacagaataatgctctgggaagaagaattccattatg	3893
Db	1735	GAATAACAAAGAACTTTTCAGCAGACATTTAGAACTGGATGCTTCAACCTCTCTGCTGTTCTT	1794
Qy	3894	gaggaagagagcaccacccagcagtcagtcactactagaattgcttgggtccttaaacgg	3953
Db	1795	GATCA-----TAGCCCTGTGGCTTCAGATATGACACTTACAGCACTGGGTCACACACA	1848
Qy	3954	ggtccactgttactgaaaaggaggtgctgactgactccttgcctccagaagaacagag	4013
Db	1849	ACTCAGGTTTCTGAAACAAAGACAAATTCGTTACATGTATATTTGTCAGAGAGGACAGAA	1908
Qy	4014	gtgaaataagaaaaaatgcccattggtattatcgcgcctgtgtccagaataatctaccgctta	4073
Db	1909	GTTAAAGTGGAAAGCGGCAATGGTCTTGGCAGCAITTTGTCAGAGATCAACTGTATTA	1968

QY 4074 accagacaggggaaacccataagaactctcaggagaaagccctagaccacacttttcagt 4133
Db 1969 TCAAAAAACAGAGTAATTTATCAAGATCCAGAAAAATAT---GATCCATTAATTCATG 2025
QY 4134 gatccagacttgccataggaacttatacaggaagctgtggtcatgaaatgaacgcagtg 4193
Db 2026 CACCCGTGATCTGCTGTGGAAACACACACTAGTACGTGTGGGCACATTATGATGCCAT 2085
QY 4194 tgcctgcagaagtattttgaaagctgtacagctgagctctcagcagcgcatcttgtagc 4253
Db 2086 TGTGTGCAAGAGTATTTGATTCCTGCTCAAGCTAAGAAACAGCAAGGCAACAGAGATTA 2145
QY 4254 cttttt-----tgacttgaaagtggagaatatcttttgccctcttttgcaaatct 4301
Db 2146 CGCTTACATACAGAGTATGATGTAGAAACGGAGATTCCTTTGGCCCCCTTTGTGAATG 2205
QY 4302 ctgtgcaactgtgatcccccattctcttgcaacctcaagaataaacaagtgaagt 4361
Db 2206 TTGAGTAATACTGTTATTC-----TCGTGCTTCCTCCAGAAATATTTTACAAAC 2259
QY 4362 gcagatgctctgtccaacttttgacctggcacggtggatcacagactgttctggccaga 4421
Db 2260 AGGTTAAATTTTCAGACCAACCAATCTGACTCACTAGTGCATTAGAACAAATATCTCAGCA 2319
QY 4422 atatacaggttataataagaacatgctaagaagagaaacccaaatctcttcttctaat 4481
Db 2320 ATAAAGCATTAACAGTTTCTTAGGAAGAAAGAAAGTACTCTCTAATAATGCCCTACAAAG 2379
QY 4482 caaggaatgggagattctacttggagttccattccatcctgagtttggcgttgagtc 4541
Db 2380 AATTACAGAAATGTGGATGAATACAGCTCCCTGGAAGGTTTCAGGCCTGATTTTCGTCT 2439
QY 4542 tcatataatattcaaatagcatcaaggaatgggtattctctcttgcgcaacaatttat 4601
Db 2440 AAGATCCCTTATCTCAGAGCATAAAGAAATGCTAAACGACATTTGGAAGTGTACTCTAC 2499
QY 4602 agaattgattgaagtgccacctgatgaaggggatactcctcgagtcgccatgctgaacctg 4661
Db 2500 AAGGTGGGACTAAAGGTTTCATCCCAAGAGAGATCCCTCGTGTGCCCAATAATGTGTGG 2559
QY 4662 agcactcgcttctcactatccaggcaattgaaatctattggagatgaagaaacact 4721
Db 2560 GTAGCTGGCGTACACCATCCAAAGCATAGAAAGAAATTTTGAGTGATGAAGATAAACCA 2619
QY 4722 ctgtttgagcacttcaaaataggcagcataatgtctgaaagcattaaatgagtttga 4781
Db 2620 TTGTTTGGTCTTACCTTGACAGCTGGATGACTGTCTTAGTCAATTGACGAGATTTGCC 2679
QY 4782 gtgcacagagattacctgtcctcaggtcctgtatagacaaacatctggtcttctcta 4841
Db 2680 GCAGCACACTGGACAGTGGCATTCAGTTTCAGTGGTGCAAGGACATTTTGTAAACTTTT 2739
QY 4842 tcagttgtctctcctaataataacacagagatacaccatgctctctgtctatagatctg 4901
Db 2740 GCATCACTGGTCCCTAATGACAGCCATGAGGAACCTTCCATGATATATGACATG 2799
QY 4902 ttctatgttttgggtggtgctgtgttagcattcccatctctgtatttgggtagccctgtt 4961
Db 2800 TTTCAATTTATTTGGTGGCTGTGTGCTTGCAATTTCTCGCTTGCAAGT----- 2846
QY 4962 gatctgcagcctctctcctgtagtttctctataaccacctttatctctcctattgac 5021
Db 2847 --TCAGGATTTTTCAGGGATCAGCCCTGGCACTGGGATTCCTTCCATTTTCCATCTGGTT 2904
QY 5022 accatggcacatgctctcagataactactacagtagacacagccc---tacccttctgt 5078
Db 2905 ACTATGGCACACATCATCAGATCTTTACTTACTCTATGACAGAGAAATGCGATGGAT 2964
QY 5079 caggttcaagaagacagtgagaggtctcattccgcatctcttctcttccttcagaaattct 5138
Db 2965 CAAGAAATATCCCTTGTGCAAGAAATCAGCAGTCTCTGCTTTGTATAAACACTTCAC 3024
QY 5139 caataacaagtggtccctctggtgtgatatctcctgctggtatttgggtctcactg 5198

Db 3025 CAGTATACGGGAAGTGCCCTTGAAAGAAATACCATCCGGCTGGCATCTGTGGAGGAGTGC 3084
QY 5199 aagaatggcatcaccccttatctctgctgtgctgcatgttttttccactatttacttggg 5258
Db 3085 AGACCTGGAATCATGCTTCTCTGAAGTGTCTGCTTATTTTTCATTACTTAATGGA 3144
QY 5259 gtaactccgcctgaggaaactgcataccaaattctgcagaagagagtagcagtgctactgt 5318
Db 3145 GTTCTCTCCACCGACATTCACATTCACATTCCTTTTCAAGAAATAGTCAGATAATGT 3201
QY 5319 agctatctatcttacctacaaatttggctctctcctccaggaatatggatactgta 5378
Db 3202 AGCTATCTTCTCCCTACCAACAACTCAATTCATTTGCTCTTTTCAAGAAATAGTCAGATAATGT 3261
QY 5379 aggccttctgctccagaggtggtgcagatcctccttactaaactgtttggaagcaaaa 5438
Db 3262 AATTCACGTGATGAAGTGTGCTGCGTAACAGTGAAGTAAAGATATCTAGAAGTGAA 3321
QY 5439 aacacctggtgcaggtaccctagaaaaaagaatagttgatagagcttctctgactat 5498
Db 3322 AGAGATGCTATAAGATATCCAAGAGAAATCTAACAATTAATAAACCCTCCAGAGGATTAC 3381
QY 5499 agctgctcctgaaatcaagcttctcattcaggtgcccacggtctgcagatgtagcga 5558
Db 3382 AGCAGCTCATTAATCAAGCATCCAAATTTCTGTCGCCGAAATCAGGTGGTGATAAGAGC 3441
QY 5559 aagcactcctgctcctgctcttctgtggtgctatactatgtctcagaacaatttgcgc 5618
Db 3442 AGACCCCAACTCTGTCCTTGTGTGCGGATCTCTGCTGTGCTCCAGAGTACTGCTGC 3501
QY 5619 caggaaatttgaaagggaaggttgaggttgcatttttccacgacttccactgtgga 5678
Db 3502 CAGACTGAACCTGGAAGGGAGGATGTAGAGCCTGCACAGCTCACACCTACTCTGTGCGC 3561
QY 5679 gcggagctgcatcttctcctaaataacagaaatgcagagtgccagtggtccttggtgaaggtaaa 5738
Db 3562 TCTGGAGTGGGCATCTTCTCTGAGAGTACGGGAATGTCAGGTGCTATTTTAGCTGGCAAA 3621
QY 5739 gcagagagctgctcctcctcctccttacttgcattggtgagatgaaatggagaacacacctggc 5798
Db 3622 ACCAAGGCTGTTTTTATCTCTCTCTTACCTTGATGACTATGGGAGACCCAGCAGGA 3681
QY 5799 ctgaagaggggcaacccctcatttctcgtgagcgggtatcggaagctccatttggct 5858
Db 3682 CTCAGACGGGAAATCTTTTACATTTATGCAAGAGCGATTCAAGAAGATTCAAGAAGCTC 3741
QY 5859 tggcaacaacactgcattatagaagagattgctaggagcccaagagactaaatcagatgta 5918
Db 3742 TGGCACCACACACAGTGTACAGAGGAAATGGACATGCACAGGAAGCAATCAGACACTG 3801
QY 5919 ttggattcactggca 5935
Db 3802 GTTGGCATTCAGTGGCA 3818

RESULT 9

HSX521.

LOCUS HSX521 818 bp DNA PRI 04-OCT-1995

DEFINITION Human partial cDNA sequence, clone x521.;

ACCESSION 247040

VERSION 247040.1 GI:1008975

KEYWORDS human.

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 818)

AUTHORS Chianilukichai,N., Pasturaud,P., Richard,I., Auffray,C. and Beckmann,J.S.

TITLE cDNA selection in the LGMD2A region

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 818)

Genethon.
Direct Submission
Submitted (05-DEC-1994) Genethon, B.P. 60, 91002 Evry Cedex France
E-mail: Nuch@genethon.fr
3 (bases 1 to 818)
Chiannikuichai,N., Pasturaud,P., Richard,I., Auffray,C. and Beckmann,J.S.
TITLE
JOURNAL
MEDLINE
FEATURES

A primary expression map of the chromosome 15q15 region containing the recessive form of limb-girdle muscular dystrophy (LGMD2A) gene Hum. Mol. Genet. 4 (4), 717-725 (1995)

Location/Qualifiers
1..818
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/clone_lib="CEPH YAC clone 774G4 was used for cDNA selection"
/map="15q15"
/tissue_type="fetal muscles"
BASE COUNT 225 a 177 C 195 g 221 t
ORIGIN

Query Match 12.1%; Score 764.2; DB 9; Length 818;
Best Local Similarity 98.2%; Pred. No. 9.2e-182;
Matches 783; Conservative 0; Mismatches 13; Indels 1; Gaps 3;

QY 5302 agtaagtgcactctgtagctatctattcattacatacaaaattgttcctgtctttccagg 5361
Db | |||||
3 ATTCCAGTGCACCTCTGTGGTAATCTATTTCACCAAAATGGGTTCCTGCTTTCCAGG 62

QY 5362 aaattggatactgtaagggcccttgctccagagtggtgctgcagatcctgccttactaa 5421
Db | |||||
63 AATATTGGGATACTGTAAAGCCCTTGCTCCAGAGTGGTGTGCAGAGTCTCCTGCCATTACCAA 122

QY 5422 actgtttaagcaaaaaaacaccggtgctcaggtaccctagaagaaaataagttgatag 5481
Db | |||||
123 ACTGTTTTGAAGCAAAAAAACACCCTGTGTCAGGTACCCCTAGAAAAAGAAATAGTTGATAG 182

QY 5482 agcttctgtagctatagctgctcctcgatgaatcaaggttctcattcagggtgccacaggt 5541
Db | |||||
183 AGCTTCTGTATGACTATAGTGTGCTCTCGAATCAAGCTTCTCATTTTCAGSTGCCACGGT 242

QY 5542 ctgcagatgatgcgaaaagcatcctgtcctctcctgtccttttctgtggagctatactatt 5601
Db | |||||
243 CCGCAGATGATGAGCAAAACATCTCTGCTCTGCCCTTTCTGTGGGGCTATATCTATGTT 302

QY 5602 ctcgaaacatttgtctgcaggaaattgtgaacgggggaaggttggagcttgcatttttc 5661
Db | |||||
303 CTCAGAACAATCTGTCAGAGAAATGTGAACGGGGGAAGAGTTGGAGCTTGCAATTTTC 362

QY 5662 acgaacttaacttgagccggaggtctgcattttcctaataaatcagagaatgccagagtg 5721
Db | |||||
363 ACGCACCTTCACTGTGAGCGCGAGTGTGCATTTTCTTAAAATCAGAGAAATGCCGAGTGG 422

QY 5722 tcctggttaaaggtaaagccagaggtgtgcctatccagctccttacttggatgaatatg 5781
Db | |||||
423 TCCTGGTTGAAGTTAAAGCCAGAGGCTGTGCCATCCAGCTCCTTACTTGGATGAATATG 482

QY 5782 gagaaacagacctgacctgaagagggg-caaccccccttatttatctcgtgagcggtat 5840
Db | |||||
483 GAGAAACAGACCCCTGGCCTGAAGAGGGGCCAACCCCCTTCATTTATCTCGTGAGCGGAT 542

QY 5841 cggaggtccatttgtctgccaacaacactgcatlataagaagagattgcttagagccaa 5900
Db | |||||
543 CGGAAGCTCCATTGTGCTGGCAACACACTGCAATTTATAGAAGAGGTGCTTAGGAGCCA 602

QY 5901 gagactaactcagatttatttgattcaactgcagttactgtgagctccaactcgtct 5960
Db | |||||
603 GAGACTAATCAGATGTTATTGTGATTCAACTGGCAGTTACTGTGAGCTCCAACTCTGCTT 662

QY 5961 caagacaatcacaaatlacgacacagtagtaagggctgattccaaattatggaaacctttct 6020

Db	663	CAAGACAATTCACAAATGACGACAGTAGTAAGGCTGATTCCAAATATATGGAACACTTTTCT	722
Qy	6021	gaggctcggaagattgagaggtcttttggctccattgcctgcaggttcacttacatcaata	6080
Db	723	GAGGCTGCTGAAGTATTTGGAGGCTCTTTTGCCTCAATGTCAGGTTCACTTACATCAATA	782
Qy	6081	aaatatcttcaatgga	6097
Db	783	AAATATTTCTTAATGGA	799
RESULT	10		
LOCUS	AK026998	2958 bp	29-SEP-2000
DEFINITION	Homo sapiens	CDNA: FLJ23345 fis, clone HEP13679.	
ACCESSION	AK026998		
VERSION	AK026998.1	GI:10439998	
KEYWORDS	oligo capping; fis (full insert sequence).		
SOURCE	Homo sapiens	hepatoma cell_line:HepG2	cdna to mRNA, clone_lib:HEP
ORGANISM	clone:HEP13679.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	1 (sites)	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE	Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,		
AUTHORS	Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,		
TITLE	Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.		
JOURNAL	NEDO human cDNA sequencing project		
REFERENCE	unpublished (2000)		
REFERENCE	2 (bases 1 to 2958)		
AUTHORS	Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,		
TITLE	Shibahara,T., Tanaka,T. and Nakamura,Y.		
JOURNAL	Direct Submission		
REFERENCE	Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio		
AUTHORS	Sugano, Institute of Medical Science, University of Tokyo,		
TITLE	Laboratory of Genome Structure Analysis, Human Genome Center;		
JOURNAL	Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan		
REFERENCE	(E-mail:cdnal@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,		
AUTHORS	Fax:81-3-5449-5416)		
COMMENT	NEDO human cDNA sequencing project supported by Ministry of		
TITLE	International Trade and Industry of Japan; cDNA full insert		
JOURNAL	sequencing; Research Association for Biotechnology; cDNA library		
REFERENCE	construction, 5'- and 3'-end one pass sequencing; Department of		
AUTHORS	Virology and Human Genome Center, Institute of Medical Science,		
TITLE	University of Tokyo (partly supported by Science and Technology		
JOURNAL	Agency).		
FEATURES	Location/Qualifiers		
source	1..2958		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/cell_line="HepG2"		
	/cell_type="hepatoma"		
	/clone="HEP13679"		
	/clone_lib="HEP"		
	/note="cloning vector pME18SFL3"		
BASE COUNT	940 a 600 c 638 g 780 t		
ORIGIN			
Query Match	10.3%;	Score 651.6;	DB 9; Length 2958;
Best Local Similarity	54.6%;	Pred. No. 2.8e-153;	
Matches 1555;	Conservative 0;	Mismatches 1229;	Indels 66; Gaps 1
Qy	3113	agtggccacatttaagaaaccagggtgatcaggccatggagtttatgaactaaagatga	3172
Db	9	AGTTGCCATTTTCAAGAAACCTGGATTAAACAGGACGAGGCATGTATGAACCTGAACACCA	68
Qy	3173	atcaactgaagacttcaaatatgtacttttatcattactcctaaaaaccagcatgcaagc	3232
Db	69	ATGTGCGCAAGAGTCTCAACTGTATTCTTCTACTTTTCAGGCGAGAACATGCCAAGC	128
Qy	3233	taacatatgcgaacaaagagagaaacaaagaaacaaacataaagcatttcgccacc	3292


```

Qy 6282 ctctcgcaatttgatttggtgttt 6308
Db 41342 CTCTGCAATTTTGTATTGGTGTTT 41316

RESULT 12
AK026948 2512 bp mRNA PRI 29-SEP-2000
LOCUS Homo sapiens cDNA: FLJ23295 fis, clone HEP10637.
DEFINITION AK026948
ACCESSION AK026948
VERSION 1 GI:10439928
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens hepatoma cell_line:HepG2 cDNA to mRNA, clone_lib:HEP
clone:HEP10637.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (sites)
AUTHORS Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,
Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
NEDO human cDNA sequencing project
Unpublished (2000)
2 (bases 1 to 2512)
AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
Shibahara,T., Tanaka,T. and Nakamura,Y.
Direct Submission
TITLE Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio
JOURNAL Sugano, Institute of Medical Science, University of Tokyo,
Laboratory of Genome Structure Analysis, Human Genome Center;
Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
[E-mail:cdnal@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416]
COMMENT NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing; Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing; Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).

FEATURES
source 1..2512
location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="HepG2"
/cell_type="hepatoma"
/clone="HEP10637"
/clone_lib="HEP"
/note="cloning vector pME18SFL3"
BASE COUNT 785 a 511 c 543 g 673 t
ORIGIN

Query Match 7.7%; Score 487.8; DB 9; Length 2512;
Best Local Similarity 54.1%; Pred. No. 6.1e-112;
Matches 1220; Conservative 0; Mismatches 987; Indels 50; Gaps 9;

Qy 3694 caggatcggaatctattaaagaatgatgagattactatgataaagaagaacgaacgaa 3753
Db 150 CAGAGACAGAGAGACCATTAATGGAAGAGAGTTCAAGGGACAGAGACAAAGCTCAGAGGA 209
Qy 3754 aaagaaagctgaagctgtaggtacatgcgccagaagatcaggtcagatgctgcct 3813
Db 210 AGAGAAAAGCAGAGATTGCCAGACTGGCCAGAGAAAAGATCATGGCTCAGATGCTGAAA 269
Qy 3814 tacagaaaaacttcattgaactataaactcatgtatgatacatacagaataatgctg 3873
Db 270 TGCAGCGGCATTTTATTGTGTAACAAACAAAGAACTCTTTTCAGACAGACATTAGAACTGGATG 329
Qy 3874 ggaagaagattcattatgaggaagagagaccccgagcagtcagtcactctagaa 3933
Db 330 CCTCAACCTCTGCTGTTCTTGATCA-----TACCCCTGGCTTCAGATATGACACTTA 383

```

```

Qy 3934 ttgcttgggtctctaaacgggggtccatctgttacttgaaaagagggtgctgacgtgatcc 3993
Db 384 CAGCACTGGTCCCCACACAAACTCAGGTTCCCTCAACAAAGACAATTCGTTCATATAT 443
Qy 3994 ttgccaagaagaacagagggtgaaaaatagaataaataatgccatggtattatcgccctgtg 4053
Db 444 TGTGTCAGAGGAGCAGAGAGTTAAAGTGAAGAGGAGGCAATGGTCTTGGCAGCAATTTG 503
Qy 4054 tccagaaatctactccttaacccagcagcaggggaaaccccatagaaactctcaggagaag 4113
Db 504 TTCAGAGATCAACTGTATTATCAAAAAACAGAAGTAATTTATTCAAGATCCAGAAAAA- 562
Qy 4114 ccctagaccacttttcattgatccagacttgccatataagaactatacagaagaagctgtg 4173
Db 563 --TATGATCCCATTTATTCATGCACCCCTGA--TCTGCTTGTGGAACACACACTAGTA-CTGTG 618
Qy 4174 gtcattgaatgcacagctgctgctgcagagaatttttgaagctgtacagctgagctctc 4233
Db 619 GGCACATTATGCATGCCCATTTGTTGGCAAGGTATTTTGTATTCCTGTTCAAGCTAAAGAAC 678
Qy 4234 agcagcgcatctcatgttgacctttt-----tgacttggaaagtggagaatc 4281
Db 679 AGCGAAGGCCAACAGAGATTACGCTTACATACGAGCTATGATGTAGAAAAACGGAGAATTCC 738
Qy 4282 ttggccctctttgcaaatctctgtgcaatactgtgatacccatcttcccttgcacctc 4341
Db 739 TTTGCCCTCTTTGTGAATGCTTTGAGTAATCTGAGTACTGTTATTCC-----TCTGCTCTCCCTC 792
Qy 4342 aaagataaacagtgagaatgcagatgctctgtcctcaacttttgacctggtgcagctgga 4401
Db 793 CAAGAAATATTTTAAACACACAGGTTAAATTTTTCAGACCAACCAATCTGACTCAGTGGGA 852
Qy 4402 tacagatgttcttgccagaatatcaggttataataataagacatgctaaagagagaaccc 4461
Db 853 TTAGAACAAATATCTCAGCAAAATAAAGCATTAACAGTTCTTAGGAAAGAAAGAAAGTACTC 912
Qy 4462 caattctctatttcttaatacagaagatgggagattcttacttggagttccatccatcc 4521
Db 913 CTAATAATGCCTCTACAAAGAAATTCAGAAAAATGTGGATGAATTACAGCTCCCTGGAAGGT 972
Qy 4522 tgagtttggcggtgagctctcgattaaattatcaaaatagcatcaaggaatggttattc 4581
Db 973 TCAGGCTTGATTTTCGTCCTAAGATCCCTTATCTGAGAGCATAAAGAAATGCTTAACGA 1032
Qy 4582 tcttggccacaacaaatttataagaattgattgaagtgccacctgatgaaaggatccctc 4641
Db 1033 CATTTGGAACCTGCTACCTACAGGTGGGACTAAAGTTTCATCCCAATGAAGAGTCTCTC 1092
Qy 4642 gagtcccatgctgacctggagcactgogcttccactatccaggcaattgaaaatctat 4701
Db 1093 GTGTCCCATTAATGTGTTGGGTAGCTGCGGTACACCATCCAAAGCATAGAAAGAAATTT 1152
Qy 4702 tggagatgaagaaacacctctgttggagcacttcaaaataggcagcataatggtctga 4761
Db 1153 TGAGTGATGAAGATAAACCAATTTGTTGGTCCCTTACCTTGCAGACTGGATGACTGCTTTA 1212
Qy 4762 aagcattaatgcagtttgctgagtgacagaggaatcacctgtcctcaggtccctgatacaga 4821
Db 1213 GGTCAATTACAGAGATTTGCCCGCAGCAGACTGGACAGTGGCATCAGTTTCAGTGGTGCAG 1272
Qy 4822 aacatctggttctgtcttctatacagttgttcttcttaacataaaaatcagaagatacaccat 4881
Db 1273 GACATTTTGTGTAACATTTTTCATCTGCTGCTGCTTAAATGACAGCCATGAGGAATTCAT 1332
Qy 4882 gccttctgtctatagatctgtttctaatgttttggtgggtgctgtgttagcatcccatcct 4941
Db 1333 GCATATTAGATATTGACATGTTTCATTTATTGTTGGGGCTTGGTGTGCTGCTGCTGCTGCT 1392
Qy 4942 tgattgggagaccctgttgatctgcagcctcttctcagttagttcttctcctaaccacc 5001
Db 1393 TGCAGTG-----TCAGGATTTTTCAGGATACAGCCTTGGCACTGGAGACC 1437
Qy 5002 ttatctcttccatttgatcacatggcacacatgcttctcagatactactactacagtagaca 5061

```



```
Db 1438 TTCACATTTCCATCTGGTTACTATGCGACACATCATACAGATCTTACTTACCTCATGTA 1497
Qy 5062 c---aggccctacccttgctcaggttcaagaagacagctgaagagagctcattccgcattctt 5118
Db 1498 CAGAGAGATGGCATGATGATCAAGAAATCCCCCTTGTGAAGAGATCAACAGCTTCTTG 1557
Qy 5119 cttctttgcagaattctcattatatacaagtggtccattgggtgtgatatctcgtgct 5178
Db 1558 CTTTGTATAAACACTTCCACAGTATACGGGAAGTGCTTGAAGAAATACCATCCGGCT 1617
Qy 5179 ggtatttggtggtcactcaagaagaatggaatcacccttattctcgtgtgctgctcattgt 5238
Db 1618 GGCATCTGTGGAGAGGTGTCAGAGCTGGAATCATGCTTTCCTCGAAGTGTCTCTCTTAT 1677
Qy 5239 ttttccactatttacttgggtgaactccgcctgaggaactgcataaccaattctgcagaag 5298
Db 1678 TTTTTCATTAATAATGGAGTCTCTCCACCCGACATCAAGT---TCCTGGACAA 1734
Qy 5299 gagagtcagtcgactctgactatctatctatttaccatacaatttgcctgctcttcc 5358
Db 1735 GCCATTTGAACATTTATGTAGCTATCTTCCCTACCACAAACACTCATTTGCCCTTTTC 1794
Qy 5359 aggaatttgggtactgttaagcccttgcctccagaggtggtgctgagatccttcccttac 5418
Db 1795 AAGAAATAGTAGATAATGAATTCATCTGATTTGAAGTGTGGTCCGTAACAGTGAAGTTA 1854
Qy 5419 taaactttgaagcaaaaaaacaccgtggtcaggtaccctagaaaaagaataagtltga 5478
Db 1855 AAGATATCTAGAGGTGAAGAGATGCTATAGATATATCAAGAGATCTAACAAATTA 1914
Qy 5479 tagagctctctgactgactatgctcctctgaaatcagcttctcattcaggtgcccac 5538
Db 1915 TAAACCTTCAGAGGATTCAGAGCTCATTAATCAAGATCCAAATTTCTGTCGCCGA 1974
Qy 5539 ggtctgcagatgatgagcgaagacatctgctcctctctcttctgtgtgggctatactat 5598
Db 1975 AATCAGGTGCTGATAGAGCAGAGAGCCCAACTCTGTGCTGTGTCGGGATCTCTGCTGT 2034
Qy 5599 gttctgaacatttgcctccaggaataattgtgaacggggaagagttggagctgcaattt 5658
Db 2035 GCTCCCAAGATTACTGCTGCGCACTGAATGGAAGGGGAGGATGTAGAGCCCTGCACAG 2094
Qy 5659 ttcaagcaacttcaactgtggagcggagctgtgctgtcttctccataaaatcagagaatgccag 5718
Db 2095 CTCACACTACTCTCTGTGGCTCTGGAGTGGCATCTTCTCTGAGAGTACGGGAATGTGAGG 2154
Qy 5719 tggctcgtggtgaaggtaaagccagagggctgtgctcctatccagctccttacttggatgaat 5778
Db 2155 TGCTATTTTGTAGTGGCAAAACCAAGGCTGTTTTTATTTCTCCCTTACCTTGTGATGACT 2214
Qy 5779 atgagaaacagaccctggtcctgaagaggggcaaccccttcttctgtgagcgggt 5838
Db 2215 ATGGGGAGACCCAGGAGGACTCAGACGGGGAATCTTTACATTTATGCAAGAGCGCAT 2274
Qy 5839 atcggagctcccttggctgtgcgaacacactgcattatagaagagatgctgtaggagcc 5898
Db 2275 TCANGAGATTCAGAGCTCTGGCACCACACAGTGTTCACAGAGAAATGGACATGCAC 2334
Qy 5899 aagagactaatcagatgttatttgattcaactggca 5935
Db 2335 AGGAAGCAATCAGACTGTTGGCATTCATGCTGGCA 2371

RESULT 13
AC068724/c
LOCUS
DEFINITION Homo sapiens' chromosome 15 clone RP11-473C18 map 15q15, complete
sequence.
ACCESSION AC068724
VERSION AC068724.7
KEYWORDS GI:12739735
SOURCE human.
```

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 190727)

Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,

Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G., Kaur,A., Madan,A.,

Nesbitt,R., Traicoff,R. and Hood,L.

Sequencing of human chromosome 15 D15S146-D15S117 region

unpublished

2 (bases 1 to 190727)

Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B.,

Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G., Kaur,A.,

Madan,A., Nesbitt,R., Shaffer,T. and Hood,L.

Direct Submission

Submitted (08-MAY-2000) Multimegabase Sequencing Center, Institute

for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA

98105, USA

3 (bases 1 to 190727)

Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,

Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G., Kaur,A., Madan,A.,

Nesbitt,R., Traicoff,R. and Hood,L.

Direct Submission

Submitted (10-FEB-2001) Multimegabase Sequencing Center, Institute

for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA

98105, USA

On Feb 10, 2001 this sequence version replaced gi:11138174.

----- Genome Center

Center: Multimegabase Sequencing Center

Web site: http://chroma.mbt.washington.edu/msg_www

Contact: leerowen@systemsbiology.org

----- Summary Statistics

Sequencing vector: pUC18; L08752

Chemistry: Dye-terminator Big Dye; 90% of reads

Assembly program: Phrap; version 0.990399

Note: Data from overlapping BACs AC009825 [Drafting center: WIBR]

and AC009852 [Drafting center: UWMSC] were added for finishing

Location/Qualifiers

1..190727

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="15"

/map="15q15"

/clone="RP11-473C18"

/clone_lib="RPC1 human BAC library 11"

/note="This clone overlaps RP11-2016 AC021899, RP11-19D14

AC009825 and RP11-402F9 AC009852. Data from AC009825 and

AC009852 were added and the consensus sequence determined

from RP11-473C18 to the extent possible"

1..9704

/note="overlap with RP11-2016, AC021899"

42130..42140

/note="low quality data"

43640..43830

/note="low quality data"

69905..70000

/note="single read coverage"

87320..190727

/note="overlap with RP11-19D14, AC009825"

99800..99900

/note="low quality data"

103370..103380

/note="low quality data"

105748..190727

/note="overlap with RP11-402F9, AC009852"

BASE COUNT 54575 a 40945 c 40610 g 54597 t

ORIGIN

Query Match 5.6%; Score 354; DB 9; Length 190727;

Best Local Similarity 100.0%; Pred. No. 5 6e-78;

Matches 354; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 423 atgcaacagttccagcagcaccacacagatatacaacattcagctaccagagctacgc 482
|||||

Db 45831 ATGCAACAGTTCCTCCAGCAGCACCACACAGATATACAACCATTCAGCTACCAAGAGCTACGC 45772
|||||

QY 483 ctgataaattagagaggggaaaaaaattctccagtccttcagtcgtagcgttgcctc 542
|||||

Db 45771 CTGATAAATTAGAGGGGAAAAAAATCTCCAGTCCCTTCACGCTGCTGACGCTTGCTTC 45712
|||||

QY 543 cggaaagcggcggcgaagcactctcagctcgtcgtcaaacccacttcagggccctc 602
|||||

Db 45711 CCGGAAGCGGGCCGGAAGCACTCTCTGAGTCTGCTCAACCCGACATTCAGGGGCCGTC 45652
|||||

QY 603 gtaaaagtgtcgtcctgtctctccgaccgcccacaggtttccctgcttgccttgccctg 662
|||||

Db 45651 GFAAAGTGTCTGCTCTCTCTCCGACCGGCCACAGGTTCTTCGCTGCTGCGCCGG 45592
|||||

QY 663 ggtcgcaactgcagcgtcagttccctcagatggtcggaagagagctgaggtact 722
|||||

Db 45591 GGTGCGCAACTGCAGCGCTGAGTTTCCCTCAAGATGGCGGACGAGGAGCTGGAGGTACT 45532
|||||

QY 723 gagaagatggaatcagcagcagaggttaccacagaccctcagcgtcgcctatc 776
|||||

Db 45531 GAGAGGATGGAATCAGCGCGGAGTTACCCAGACCCCTCAGCGTGTGCATCT 45478
|||||

RESULT 14

AC090992

LOCUS AC090992 61896 bp DNA HTG 22-MAR-2001

DEFINITION Homo sapiens chromosome 15 clone RP11-473C18 map 15, LOW-PASS

SEQUENCE SAMPLING.

AC090992

VERSION AC090992.1 GI:13431042

KEYWORDS HTG; HTGS_PHASE0.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 61896)

AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,

Barnas,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B., Brown,A.,

Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,

Collins,S., Collymore,A., Cooke,P., DeArelano,K., Dewar,K.,

Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D.,

Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,

Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,

Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRocque,K.,

Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,

McClean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,

McEwan,P., McKernan,K., McPheeters,R., Meldrum,J., Meneus,L.,

Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,

Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,

Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,

Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,

Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,

Severy,P., Sougnuez,C., Spencer,B., Stange-Thomann,N.,

Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,

Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,

Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,

Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

TITLE Submitted (22-MAR-2001) Whitehead Institute/MIT Center for Genome

RESEARCH, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L12436
Center clone name: 473_C_18

* NOTE: This record contains 76 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

* 1 716: contig of 716 bp in length
* 717 816: gap of 100 bp
* 817 1540: contig of 724 bp in length
* 1541 1640: gap of 100 bp
* 1641 2352: contig of 712 bp in length
* 2353 2452: gap of 100 bp
* 2453 3172: contig of 720 bp in length
* 3173 3272: gap of 100 bp
* 3273 3996: contig of 724 bp in length
* 3997 4096: gap of 100 bp
* 4097 4805: contig of 709 bp in length
* 4806 4905: gap of 100 bp
* 4906 5633: contig of 728 bp in length
* 5634 5733: gap of 100 bp
* 5734 6441: contig of 708 bp in length
* 6442 6541: gap of 100 bp
* 6542 7240: contig of 699 bp in length
* 7241 7340: gap of 100 bp
* 7341 8065: contig of 725 bp in length
* 8066 8165: gap of 100 bp
* 8166 8845: contig of 680 bp in length
* 8846 8945: gap of 100 bp
* 8946 9638: contig of 693 bp in length
* 9639 9738: gap of 100 bp
* 9739 10463: contig of 725 bp in length
* 10464 10563: gap of 100 bp
* 10564 11276: contig of 713 bp in length
* 11277 11376: gap of 100 bp
* 11377 12086: contig of 710 bp in length
* 12087 12186: gap of 100 bp
* 12187 12901: contig of 715 bp in length
* 12902 13001: gap of 100 bp
* 13002 13695: contig of 694 bp in length
* 13696 13795: gap of 100 bp
* 13796 14509: contig of 714 bp in length
* 14510 14609: gap of 100 bp
* 14610 15328: contig of 719 bp in length
* 15329 15428: gap of 100 bp
* 15429 16140: contig of 712 bp in length
* 16141 16240: gap of 100 bp
* 16241 16955: contig of 715 bp in length
* 16956 17055: gap of 100 bp
* 17056 17773: contig of 718 bp in length
* 17774 17873: gap of 100 bp
* 17874 18561: contig of 688 bp in length
* 18562 18661: gap of 100 bp
* 18662 19396: contig of 735 bp in length
* 19397 19496: gap of 100 bp
* 19497 20214: contig of 718 bp in length
* 20215 20314: gap of 100 bp
* 20315 21031: contig of 717 bp in length
* 21032 21131: gap of 100 bp
* 21132 21845: contig of 714 bp in length
* 21846 21945: gap of 100 bp

```
* 21946 22651: contig of 706 bp in length
* 22652 22751: gap of 100 bp
* 22752 23448: contig of 697 bp in length
* 23449 23548: gap of 100 bp
* 23549 24282: contig of 734 bp in length
* 24283 24382: gap of 100 bp
* 24383 25103: contig of 721 bp in length
* 25104 25203: gap of 100 bp
* 25204 25942: contig of 739 bp in length
* 25943 26042: gap of 100 bp
* 26043 26758: contig of 716 bp in length
* 26759 26858: gap of 100 bp
* 26859 27573: contig of 715 bp in length
* 27574 27673: gap of 100 bp
* 27674 28392: contig of 719 bp in length
* 28393 28492: gap of 100 bp
* 28493 29213: contig of 721 bp in length
* 29214 29313: gap of 100 bp
* 29314 30032: contig of 719 bp in length
* 30033 30132: gap of 100 bp
* 30133 30847: contig of 715 bp in length
* 30848 30947: gap of 100 bp
* 30948 31656: contig of 709 bp in length
* 31657 31756: gap of 100 bp
* 31757 32472: contig of 716 bp in length
* 32473 32572: gap of 100 bp
* 32573 33281: contig of 709 bp in length
* 33282 33381: gap of 100 bp
* 33382 34100: contig of 719 bp in length
* 34101 34200: gap of 100 bp
* 34201 34912: contig of 712 bp in length
* 34913 35012: gap of 100 bp
* 35013 35727: contig of 715 bp in length
* 35728 35827: gap of 100 bp
* 35828 36547: contig of 720 bp in length
* 36548 36647: gap of 100 bp
* 36648 37368: contig of 721 bp in length
* 37369 37468: gap of 100 bp
* 37469 38180: contig of 712 bp in length
* 38181 38280: gap of 100 bp
* 38281 39000: contig of 720 bp in length
* 39001 39100: gap of 100 bp
* 39101 39813: contig of 713 bp in length
* 39814 39913: gap of 100 bp
* 39914 40639: contig of 726 bp in length
* 40640 40739: gap of 100 bp
* 40740 41445: contig of 706 bp in length
* 41446 41545: gap of 100 bp
* 41546 42289: contig of 744 bp in length
* 42290 42389: gap of 100 bp
* 42390 43114: contig of 725 bp in length
* 43115 43214: gap of 100 bp
* 43215 43927: contig of 713 bp in length
* 43928 44027: gap of 100 bp
* 44028 44741: contig of 714 bp in length
* 44742 44841: gap of 100 bp
* 44842 45541: contig of 700 bp in length
* 45542 45641: gap of 100 bp
* 45642 46350: contig of 709 bp in length
* 46351 46450: gap of 100 bp
* 46451 47172: contig of 722 bp in length
* 47173 47272: gap of 100 bp
* 47273 47976: contig of 704 bp in length
* 47977 48076: gap of 100 bp
* 48077 48791: contig of 715 bp in length
* 48792 48891: gap of 100 bp
* 48892 49626: contig of 735 bp in length
* 49627 49726: gap of 100 bp
* 49727 50440: contig of 714 bp in length
* 50441 50540: gap of 100 bp
* 50541 51262: contig of 722 bp in length
* 51263 51362: gap of 100 bp
* 51363 52085: contig of 723 bp in length
```

```
* 52086 52185: gap of 100 bp
* 52186 52913: contig of 728 bp in length
* 52914 53013: gap of 100 bp
* 53014 53737: contig of 724 bp in length
* 53738 53837: gap of 100 bp
* 53838 54541: contig of 704 bp in length
* 54542 54641: gap of 100 bp
* 54642 55352: contig of 711 bp in length
* 55353 55452: gap of 100 bp
* 55453 56172: contig of 720 bp in length
* 56173 56272: gap of 100 bp

Query Match 4.4%; Score 279.4; DB 2; Length 61896;
Best Local Similarity 73.3%; Pred. No. 3.4e-59;
Matches 294; Conservative 0; Mismatches 106; Indels 1; Gaps 1;

Qy 423 atgcacaggtcccgagcaccacacccagatatacacaccattcagctaccagaagctacgc 482
Db 47681 ATGCAACAGTTCCAGCACCACCAACCCAGATATACAAACATTACAGCTACCAAGAGCTACGC 47740

Qy 483 ctgataaattagagggaaaaaaatctccagtccttcacgtcccttcacgtcgtacgtctgttc 542
Db 47741 CTGATAAATTAGAGGGGAAAAAAATCTCCAGTCCCTTCAGCTCCCTGACGCTTGCTTC 47800

Qy 543 cgggaagcggcgggaagccactcctcagctcgtcgtcaaaccccgacttcagggcgctc 602
Db 47801 CGGGAGCGGGCGGAAGCCACTCTCGAGTCTGCGTCAAAACCCGACTTCAGGGGCGGTC 47860

Qy 603 gtaaaagtgtctccctgtctctcggacggccacagggttcctcgtctgctctggtggcg 661
Db 47861 GTAAAGTGTCTCGCTGCTCTCCAGCGGCGCACAGGTTTCGGCTTGCGCTTGCGCGGN 47920

Qy 662 gggtcggcaactgcaggctcagtttccctcaagatggcggacgagaggtcagaggtac 721
Db 47921 GGGTCGGCAACTGCAGGCGCTCAGTTTCCTTCAGATGGCGGAGAGAGCTGGACNNNN 47980

Qy 722 tgagagggatggaatcagcgcggaggttacccagacccctcagcgtctgcatcttggtg 781
Db 47981 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 48040

Qy 782 ggatcagcaagtgtattttatctactgcttcttgcacatt 822
Db 48041 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 48081

RESULT 15
MMIUBR2
LOCUS MMIUBR2 910 bp DNA ROD 18-JUN-1998
DEFINITION Mus musculus ubiquitin-protein ligase E3-alpha (Ubr1) gene, exon 2.
ACCESSION AF067372
VERSION AF067372.1 GI:3236421
KEYWORDS
SEGMENT 2 of 9
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 910)
AUTHORS Kwon,Y.T., Reiss,Y., Fried,V.A., Hershko,A., Yoon,J.K., Gonda,D.K.,
Sangan,P., Copeland,N.G., Jenkins,N.A. and Varshavsky,A.
TITLE The mouse and human genes encoding the recognition component of the
N-end rule pathway
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 910)
AUTHORS Kwon,Y.T., Reiss,Y., Fried,V.A., Hershko,A., Yoon,J.K., Gonda,D.K.,
Sangan,P., Copeland,N.G., Jenkins,N.A. and Varshavsky,A.
TITLE Direct Submission
JOURNAL Submitted (21-MAY-1998) Division of Biology, California Institute
of Technology, 1200 E. California Bl., Pasadena, CA 91125, USA
FEATURES
source
Location/Qualifiers
1..910
/organism="Mus musculus"
/strain="129/SvJ"
```

/db_xref="taxon:10090"
/cell_type="embryonal stem"
/dev_stage="embryo"
/chromosome="2"
143. .399
/gene="Ubr1"
/number=2

BASE COUNT 273 a 150 c 168 g 319 t
ORIGIN

Query Match 3.4%; Score 215.2; DB 10; Length 910;
Best Local Similarity 89.2%; Pred. No. 3.6e-43;
Matches 232; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 777 tgggtggatcagcaagttgattttatatactgcttcttgcacatcattggcacaaattggtg 836
|||||
Db 143 TGGTGGGATCAGCAAGTTCATTTCTATATACTGCTTTCTTACATCATTGGCACAAATTAGTG 202
|||||

QY 837 ccagaaaatttactttgctgaaatggaccagacttggaaaagcagagagaaaagtgtacaa 896
|||||
Db 203 CCAGAAATTTATTTGCTGAGATGGACCCAGATTTGGAAGCAGAGAGAGAGAGTGTACAG 262
|||||

QY 897 atgtcaatattcactccactgaatggtagtacttatttggagaagatccagatatcttcta 956
|||||
Db 263 ATGTCATTAATCTACTCTCTTGGAGTGGTACTTATTGGAGAGGATCCGGATATTGCTTA 322
|||||

QY 957 gaaaaattgaagcacagtgagcatttcagcttttggagggttttccaaaagtggagag 1016
|||||
Db 323 GAGAAATTAACACACAGTGGAGCGTTCAGTTCGTGGGAAGGTTTCAAAAGTGGAGAA 382
|||||

QY 1017 acaacctattcttcaggga 1036
|||||
Db 383 ACAACATATTCTCTGTAGGTA 402
|||||

Search completed: March 1, 2002, 16:16:46
Job time: 13031 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 28, 2002, 10:00:18 ; Search time 49.84 Seconds
(without alignments)
5133.032 Million cell updates/sec

Title: US-09-724-126A-2
Perfect score: 9224
Sequence: 1 MADEAGGTRMEISAEPLQ.....EIARSQETNQLFGFNWQLL 1749

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_17.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8617	93.4	1757	11 070481	O70481 mus musculus
2	3160	34.3	1275	4 015057	O15057 homo sapien
3	3151	34.2	1273	4 09NU68	Q9nu68 homo sapien
4	2278.5	24.7	1829	5 09VX91	Q9vx91 drosophila
5	1945.5	21.1	1927	5 P91133	P91133 caenorhabdi
6	1716	18.6	333	4 060708	O60708 homo sapien
7	1692	18.3	329	4 075492	O75492 homo sapien
8	606.5	6.6	1225	10 09LZ95	Q9LZ95 arabidopsis
9	598	6.5	212	4 09H578	Q9H578 homo sapien
10	549.5	6.0	1872	3 007963	Q07963 saccharomyc
11	537.5	5.8	2168	5 019330	Q19330 caenorhabdi
12	420	4.6	1456	5 09W3M3	Q9w3m3 drosophila
13	343	3.7	795	10 09LZ94	Q9LZ94 arabidopsis
14	335.5	3.6	794	10 096248	Q96248 arabidopsis
15	311	3.4	783	5 09W3M5	Q9w3m5 drosophila
16	205.5	2.2	3187	11 063714	Q63714 rattus norv
17	190.5	2.1	3259	4 014789	Q14789 homo sapien
18	188.5	2.0	2041	4 09HAT7	Q9hat7 homo sapien
19	188.5	2.0	2047	4 09UH61	Q9uh61 homo sapien

ALIGNMENTS

RESULT 1

O70481					
ID	O70481	PRELIMINARY;	PRT:	1757	AA.
AC	O70481;				
DT	01-AUG-1998	(TREMBLrel. 07, Created)			
DT	01-AUG-1998	(TREMBLrel. 07, Last sequence update)			
DT	01-JUN-2001	(TREMBLrel. 17, Last annotation update)			
DE	UBIQUITIN-PROTEIN LIGASE E3 COMPONENT N-RECOGNIN (UBIQUITIN-PROTEIN				
DE	LIGASE E3-ALPHA).				
GN	UBR1.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Kwon Y.T., Reiss Y., Fried V.A., Hershko A., Yoon J.K., Gonda D.K.,				
RA	Sangan P., Copeland N.G., Jenkins N.A., Varshavsky A.;				
RL	Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.				
RN	[2]				
RP	SEQUENCE OF 1-446 FROM N.A.				
RC	STRAIN=129/SVJ;				
RA	Kwon Y.T., Reiss Y., Fried V.A., Hershko A., Yoon J.K., Gonda D.K.,				
RA	Sangan P., Copeland N.G., Jenkins N.A., Varshavsky A.;				
RT	"The mouse and human genes encoding the recognition component of the				
RT	N-end rule pathway".				
RL	Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AF061555; AAC40165.1; -				
DR	EMBL; AF067371; AAC23678.1; JOINED.				
DR	EMBL; AF067371; AAC23678.1; JOINED.				
DR	EMBL; AF067373; AAC23678.1; JOINED.				
DR	EMBL; AF067374; AAC23678.1; JOINED.				
DR	EMBL; AF067375; AAC23678.1; JOINED.				
DR	EMBL; AF067376; AAC23678.1; JOINED.				
DR	EMBL; AF067377; AAC23678.1; JOINED.				
DR	EMBL; AF067378; AAC23678.1; JOINED.				
DR	MGD; MGI:127797; Ubr1.				
DR	InterPro; IPR000307; Ribosomal_S16.				
DR	Pfam; PF02207; zf-UBR1.1.				
DR	SMART; SM00396; Znf-UBR1; 1.				

Q14221 homo sapien
Q9bg87 bos taurus
Q15075 homo sapien
Q95949 homo sapien
Q9h450 homo sapien
Q14812 homo sapien
O60588 homo sapien
Q9esd3 mus musculu
Q9esd4 mus musculu
Q9esel mus musculu
Q9xwr0 caenorhabdi
Q9ncg0 drosophila
O01385 drosophila
Q9hcl1 homo sapien
Q9bzf9 homo sapien
O15154 homo sapien
Q9vki0 drosophila
Q9nud0 staphylococ
Q9ufi5 homo sapien
Q9v489 homo sapien
Q9#6v0 gallus gall
O74424 schizosacch
Q9av25 oryza sativ
O23332 arabidopsis
Q9ncf9 drosophila
O04650 arabidopsis

RX MEDLINE=97349984; PubMed=9205841;
RA Nagase T., Ishikawa K., Nakajima D., Ohira M., Seki N., Miyajima N.,
Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
The complete sequences of 100 new cDNA clones from brain which can
code for large proteins in vitro.";
RL DNA Res. 4:141-150(1997).
DR EMBL; AB002347; BAA20806.1; -;
DR InterPro; IPR000408; RCC1.
DR InterPro; IPR001841; Znf_ring.
DR PROSITE; PS00626; RCC1_2; UNKNOWN_1.
DR SMART; SM00184; RING; 1.
FT NON_TER 1
SQ SEQUENCE 1275 AA; 145331 MW; 101FF1F66E056066 CRC64;

Query Match 34.3%; Score 3160; DB 4; Length 1275;
Best Local Similarity 47.3%; Pred. No. 5.3e-212;
Matches 612; Conservative 248; Mismatches 393; Indels 40; Gaps 19;

QY 479 LKYLISKPTIWERLQMQLFEGRSFLKILTCMOGMEEIRROVGOHLEVPDWEAAIAI 538
DB 1 LKYLISKPTWSEDELKQKLEGFDAFLLELLKCMQMDPIRQVGOHIEMEPWEAAFTL 60

QY 539 QMQLKNILLMFQEWCADEELLLVAYKECHKAVMRCSTSFSSSKTVVQS-CGHSLKTS 597
DB 61 QMQLKTHVTSMQDWCASDEKVLIEAYKKLAVLMQCHGGYTDGEQPTLSICGHSVETR 120

QY 598 YRVSEDLVTHLPSRTLGLHVLRLGAVSRLEHFVSFDFQVEVLVEPLRCLVLVA 657
DB 121 YCVSQEKVSHLPSVRLAGLHVLRLSEVAYKPELPLSELPMPMLIEHLRCLVLCA 180

QY 658 QVAEMRRNGLSLISOVYVQDKREEMVDKDIIMLOIGASLMDPNKFLLLVLORVEL 717
DB 181 QVHAGMRNGFSLVNIQYHHNVKCRREPFDKDVMLOTGVSMMDPNHFLMILMSREL 240

QY 718 AEAFN-----KTISK--DODLIKQYNTLIEMLQVLIYIVGERYVPGVGNVTKKEEVTMR 770
DB 241 YQIFSTPDYGRFSEITHKDVQCNQNTLIEMLYLIIMLVGERFSGVGQVGNATDEIKR 300

QY 771 EIIHLICIEPMPHSAIAKNPENNETGLNVINKVATPKPGVSGHGVYELKDESLKD 830
DB 301 EIIHQLSIKPMAHSELVKSUPDENKETGMESVIEAFAHFKPLGTGRMYELKPECAKE 360

QY 831 FNMVYHYVTSQHSKABHMKRRKQENKDEALPPPPPPPCPAFAKVINLLNCIDIMYI 890
DB 361 FNLVYHFSRAEQSKAEARQKLQRNEDTALPPVLPPECPFLASVLIQSDVMLCI 420

QY 891 LRTVFERAIDTSDNLMTGMLQAFHIALGLLEEKQLOKAPEEE-VTFDFYHKASRLG 949
DB 421 MGTILQWAVEHNGYAWSESMLQRLVHLIGMALQOEKQHLENVTEHVVTFTTQKISKPG 480

QY 950 SSANW---IQMLLEKLGIFQLEGQKDMITWILQMFDTVTKRLRKS-CLIVATSGSESI 1005
DB 481 EAPKNPSILAMLETLQNAPLYEVHKDMIRWLKTFNNAVKMKRESSPTSPVAETEGTIM- 539

QY 1006 KNDIEITHDKAEKPKRAEARLHRQKIMQMSALQKNEFETHKLMYDNTSEMPCKEDSI 1065
DB 540 --ESSRDKDKAEKRAEATARUKREKIMQMSQEMQHRFIDENKELFOQTLELDASTSAV 597

QY 1066 MEEESTPAVSDYSRIALGPKRGSPSVTEKEVLTCLTCQEEQEVKIENNAMVLSACVQKSTA 1125
DB 598 LDH--SPVASDMLTALGPTQTQVPEORQFVTCILCQEEQEVKVESRAMVLAFAVQRSTV 655

QY 1126 LQHRGKPIELSGEALDFMDDPLAYVTGSGGHVMAVCWKQYFEAVOLSSQ----- 1181
DB 656 LSKNRSFIQ-DPEKYDPLEMHPDLSCGTHTSSCGHIMHAHCWQRYFDSVQAEQRQOR 714

QY 1182 -RIHVDLFDLESGEYLCPLCKSLCNTVIPPIPOPOKINSENADALAQLLTLARWIQTVL 1240
DB 715 LRLHTS-YDVENGFEFLCPLCECNSNTVIPL-LPPRNFNRLN-FSDQPNLTQWRTIS 771

QY 1241 ARISGYNIRHAKGNPIPIFFNQMGMDSTLEFHSILSFGVSESSIKYSNSIKEMVILFATT 1300

DB 772 QQIKALQFLRKEESTPNNASTKNSENVDELQQLPEGRPDPRPKIPYSESKEMLTTFGTA 831
QY 1301 IVYRIGLVKVPDDRDPRVPMLTWTSTCAFTTQAIENLLGDEGKPLFGALQNRHNGLKALMQ 1360
DB 832 TYKVLGVKVPNEEDPRVPMCMGSCAYTTQSIERIISLDEKPLFGPLPCLDDCLRLSLTR 891

QY 1361 FAVAQRITCPQVLIQKHLVRLLSVLPNIKSEDTPLCLSLDLFHLVGVAVLAFPSLYWDD 1420
DB 892 FAAAHWTVASVSVVQGHFKCLFASLVPNDSHEELPCILDDMFHLVLGLVAFPALOCQD 951

QY 1421 PVDLOPSSVSSSYNHLVFLHITMAHMLQILL---TVDTGLPLAQVOEDSEEAHSASSFF 1477
DB 952 -----FSGISLGTGDLHIEHLVTHMAHIIQILLTSCTEENGMD--QENPPCEESAVLALY 1004

QY 1478 AEISQYTSGISGCDIP-GWYLVVSLKNGTTPYLRCAALFEHYLLGYTPPELHTNSAEGE 1536
DB 1005 KTLHQYT-GSALKEIPSGWHLMKRSVRAGIMPFUKCSALFFHYLNGVPPDPDIQV-PGTSH 1062

QY 1537 YSALCSYLSLPTNLFLLFOEYWDVTRPLQRCWADPALLNCLKOKNTVVYPRKRNSLIE 1596
DB 1063 FEHLCSYLSLNNLCLFQENSEIMNSLIESWCNRSNVEVKRYLEGERDAIRYPRESNKLIN 1122

QY 1597 LPDDYSCLLNQASHFRCPRSADDERKHPVLCFCGAILCSQNICCOEIVNGEEVGACIFH 1656
DB 1123 LPEDYSLLINQASNFSCPKSGGDKSRAPTLCLVCGSLLCSSQSYCCQTELEGEDVGACTAH 1182

QY 1657 ALHCAGAGVCIKIRCEVRVVEGKARGCAYPAPYLDVEYGETDPLKRGNPLHLRSRYR 1716
DB 1183 TYSCSGVGIFLRVRECQVFLAGTKGCFYSPYLDYGETDQGLRRGNPLHLCKBRFK 1242

QY 1717 KULHVQOHCIEETARSOETNOMLFGFNOLL 1749
DB 1243 KIQLWQHQSVTETEGHAQENOTLVGIDWOHL 1275

RESULT 3
Q9NU68 PRELIMINARY; PRT; 1273 AA.
AC Q9NU68;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE DJ392M17.3 (K1AA0349 PROTEIN) (FRAGMENT).
GN DJ392M17.3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN NCBI_TaxID=9606;
RP SEQUENCE FROM N.A.
RA Collier R.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL049843; CAB75421.1; -;
DR InterPro; IPR000408; RCC1.
DR InterPro; IPR001841; Znf_ring.
DR SMART; SM00184; RING; 1.
DR PROSITE; PS00626; RCC1_2; UNKNOWN_1.
FT NON_TER 1
SQ SEQUENCE 1273 AA; 145060 MW; A43DCE0BF4C72B39 CRC64;

Query Match 34.2%; Score 3151; DB 4; Length 1273;
Best Local Similarity 47.3%; Pred. No. 2.3e-211;
Matches 610; Conservative 248; Mismatches 393; Indels 40; Gaps 19;

QY 481 YLISKPTIWTWRLMQFLGFRSFLKILTCMOGMEEIRROVGOHLEVPDWEAAIAIQM 540
DB 1 YLISKPTWSEDELKQKLEGFDAFLLELLKCMQMDPIRQVGOHIEMEPWEAAFTLQM 60

QY 541 QLKNTLLMFQEWCADEELLLVAYKECHKAVMRCSTSFSSSKTVVQS-CGHSLKTSYR 599
DB 61 KLTHVISMMDWCASDEKVLIEAYKKLAVLMQCHGGYTDGEQPTLSICGHSVETR 120

Query Match		24.7%	Score 2278.5;	DB 5;	Length 1829;
Best Local Similarity		31.6%	Pred. No. 4e-150;		
Matches 606;		Conservative 305;	Mismatches 714;	Indels 295;	Gaps 55;
Qy	5	EAGTEREISAEALPQTPQRIASWMDQVDYTAFLHLHQAQLVPEIYFAEMDPDLEK---	61		
Db	24	QAGTLDRSDIIEFLKRESKP---VFDYQTSATVRGAHPRSSRP-----PDVSSFLK 72			
Qy	62	---QEEVSQMSIFPLWYLFEGDPDICEKLKHSQ-AFQLCGRVFKSGETIYSCROCAI 117			
Db	73	CMFKESLAKIEIIDVVVEFMGLDNPSSALEKLRLLEGNTATVCGKVFKNGEPTYSYCREGV 132			
Qy	118	DPTCVLCMDFQDSVHKHNRHKMTSGTGGGFCGDPTEAWKTGFCVYNHHPGRAGTTKEN 177			
Db	133	DPTCVLCVNCFKRAHFRHKYKMTSGGGGCCDGDDEAMKQOYCELHLANRKNPIESK 192			
Qy	178	SRCLNEEVIVQARKIPSPVIKYVEMTIWEEREKELPELQIREKN-----ERYVCV 229			
Db	193	I---LTDVAFLERVEICGAILAFVCVSL---ETEPNASLOCLDGNVEGGQVDAQYCTV 245			
Qy	230	LFNDEHISYDRHYISLQALDCELAERLAQLHTTADTKGRRAVKAGAYAAOEAKEDEKSH 289			
Db	246	LYNDESHTFDQVIOTLTKIAKRAKADAMEIVAAIDREGRAVVKCDTPEECNKLKVSIEHQ 305			
Qy	290	-----SENVSOHPLHVEVLHSEINAHOKFALRLGSMWNKIMSYSSDERQIEFQACLR 341			
Db	306	MILPSTSVTARNQSLRTSVLHIGAVACQOAFQALLGNFQEFVLRYHLPFKTFSELVQR 365			
Qy	342	EEPDSENPCLISRLMLDAKLYKGARKILHELIPSSPFEMEMEYKKLFAMEYKYKLOK 401			
Db	366	KQ---ETFC-IRHILEYDVKLWKARTCWHRLLISGLMEYDNKMLIAQAEESRYATIVE 421			
Qy	402	EYISDDHRSISALTASVQMTVPTLARHLIEBQNVTSVITETLLEVLPEYLDNRKNFNF 461			
Db	422	DFISDDHDAFVSLSVQLFTVPSIAHLIAHEGIFDKLLHLYHYHVAIEKFIKNTLHF 481			
Qy	462	QG---YSQDKLGRVYATCDLYILISPTIWTETRLRMQFLEGFRFLKILTCMOGMEIR 519			
Db	482	SKNTASITFKRANWILYDURLYLSLPDVLNDLNGFLGEGCALMRVLNVMOGMESMT 541			
Qy	520	RVQGHIEVDPDWEAAIAIOMQLNILLMFQEWACACDEELLIVAYKECHKAVMRCSTSF 579			
Db	542	RTQGHMDYEPENECAPNLHKLATTISQVIDWASGDVKLLKLYKMTMALV--SNSFI 599			
Qy	580	SSSTVTVQS---CHSLETYSYRVSDSLVSIHPLSLTLAGLHVRLSLGAVSRHLFEVS 636			
Db	600	VGGEKVMQPKVADHVANCLVYDISVQPVSTHPLPSRFFAGIYL---HLGA---HD-LT 651			
Qy	637	FEDFOQEV-----LVEYPLRCLVLVAQVVAEWRRNGLSLISQVFFYQDVKCRE 685			
Db	652	YDGLQTEALSKLTPREIIEPVLCQAMTAQVAGLWRNGYTLHLHLYFYFNRVCRV 711			
Qy	686	EMYDKDIIMLQIGASLMDPNKFLLLVQRYE---LAEAFNKITSTKD--QDLIKQYNTL 739			
Db	712	EMLDRDIACLOIGASLMESEFLHVLNRNTIPWQENYWSLLSGNEMNDIIRE-ASI 770			
Qy	740	IEMLQVLLIVYGRYVPGVGNVTKTEVTREIHLHCIEPMPHSAIAKNLPENE--NNE 797			
Db	771	FDEPLELLIIVIGBRWMPGVMTEEDRLRKEIIQLLICKIPYSHSELSSRALPDGNSGSD 830			
Qy	798	TGLENVINKVATFKKP-GVSGHGYVELKDRSLKDFNMFYHYKTSQHSKAHMKOKRKKQ 856			
Db	831	NVFEVINTVAVFKPKPGADSKGYELKEHLKEFNMFYHYTKEDRSKAEELQERRKA 890			
Qy	857	ENKDEALPPPPPPCFAPKFNLLNCIDIMMYILTRVVEERAIQDTSNLTMEGLQMAFH 916			
Db	891	KKQLVCCPPMLPKLTAFTPMANILQCPVFLNCSLIMERALNAYSRSFTESHLQVLH 950			
Qy	917	ILALGLEEKQOOLAKAPEEVEFTDFYHKASRLGSSAMNIQMLLEKLGIPQLEGQKDMIT 976			
Db	951	LLGVAIQEE-----LSEHYFPFLSYFERSQEYG-----ILEKLEELARCPRLEAHYDFVL 999			

Qy	977	WILQMFDTVKRLREK-----SCLIVATTSGSESINKDEITHDKKAEKRKAEEA 1026	
Db	1000	WTIERF---KLOAKOAPSDGRAGPSC-SQOQTGGKLSLSAE---QAREERENARLA 1051	
Qy	1027	RUHQKINMAQSALOKNFTHKLMYDNTSEMPGKE-----DSIMEESTPAVSQY 1077	
Db	1052	AERRAHIMAQMAKQAKSFISNAEMFADENETRKESASTGPMDEWEDIPPEEQGAVAL 1111	
Qy	1078	SRIA-LGPKRGFSVTEKEVLTCLCOEBOEVKTIENNAMVLSACVQKSTALTQHRGPIEL 1136	
Db	1112	SKVACGLGPDPRKRYHGTDDTFKILCFENCAISRGGQLVSSAFVQTSRV----- 1160	
Qy	1137	SCEALDPLFMDPDL---AYGTVTGSGCHVMAVCWKQYF---EAVOLSSQORITHVDLFDL 1190	
Db	1161	-----TFTTNLRNSQALHISCCGHVHYSWLEYFTNEEFKELRRPHRRAALAAQ 1213	
Qy	1191	ESGYLCPCLCKSLCNTVPIIPILOPO-----KINSENADALAQL--- 1229	
Db	1214	ANVEFOCPYCRTLNSNAIIPVTETLPAFSAPPSNPSYPLDSEFVEIMSTLAIELGNVKDH 1273	
Qy	1230	-LTARWQTVLARISGYNIRHAKGENPIPIFNQMGMDSTLEFHSILSGVSESSIKYSN 1288	
Db	1274	ELTTLPSVSNIL-RLSGVYVGLAQFERSQLIKNPP-----RLHADYIEGIE----- 1319	
Qy	1289	SIKENVILLFATTIYRIGLK-VPPDERDRPVPMLTWTSCFTCAIOAIENLLGDECKPLFGAL 1347	
Db	1320	-FLKALLNTMKIQOHLKDHFAIESIEMVPIL-WDSCSYTLOALEIYLAVKEPLKAE 1377	
Qy	1348	QNRHONGKALMQ-----FAVAQRITCPOVLQIKHVLRLSVLVPNIKSEDTPLCLSIDL 1402	
Db	1378	SMRHQSCARNLVRACSRSSALEWETDPLLLPPMRSAEFSSRLDITFNQNDTSVLEWDC 1437	
Qy	1403	FHVLY---GAV-LAPPSLYWDDPVDLQPSVSSVSNHLYLFLHITMAHMLQILLFVDTG 1457	
Db	1438	FRVLVPFQFGLNLMVPEKGYKTIIP-----SGSMFDFYIMOTMFLAQLTKAVLCFDVE 1491	
Qy	1458	LPLOA-----VOEDSE 1468	
Db	1492	KEKARAEKAPNSELTQLDYIEQLPSRIRDNMTDFRRYINIPARVLOKTKOKOLVEESE 1551	
Qy	1469	E---AHS-----ASSFFAEISQYTSIGSGCIDPGWYLMVSLKNGITPYLRCAALFFHYL 1519	
Db	1552	ENQHGQTVIPECESHLALLLEY-----VQRMSSFLRCSCLFYRFL 1594	
Qy	1520	LGVTPEELHTNSAGEYSALCSYLSLPTNLFLFLFOBYWDTV-RPLLQRCWADPALLNCL 1578	
Db	1595	TDVDFPDFTPTDQPD-RFDLMQCYLGLDPLMGVYFD--METVYATMMHFSAPHIDREV 1651	
Qy	1579	KOKNTVVRYPKRNSI-IE-----LPDDYSCLLNQASHFRCPSRADDERKHPVL 1626	
Db	1652	EQR-----COPDARRSLQVEPCRLPRLKVLCDDFSDLINSVSDIFCPNNREEMKPTPTM 1707	
Qy	1627	CLFCGAILCSQNICQIEVNGEEVGACIEFHALHCGAGVCIFFKIRECRVVLVEGKARGCA 1686	
Db	1708	CLICGLILCGOSYCCOPELGLKVSVGACTHHAACGAEGVIFLRIDCQVYVL-GRGKGF 1766	
Qy	1687	YPAYLDEGETDPLKRGKNPLHLSRERYKHLVWQOHCIIIEIARSQETNMLFGFNW 1746	
Db	1767	VPYPYLDEYGETDMGLRRGNPLRLSQAAVRKYIQLWHLGHLGEIARLNDNANVAAAQW 1826	

RESULT 5
P91133
ID P91133 PRELIMINARY; PRT: 1927 AA.
AC P91133;
DT 01-MAY-1997 (Tremblrel. 03, Created)
DT 01-MAY-1997 (Tremblrel. 03, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE SIMILAR TOS. CEREBISIAE UBIQUITIN-PROTEIN LIGASE E3 COMPONENT
DE SP:P19812.
GN C32E8.11.
OS Caenorhabditis elegans.

Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W., Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K., Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J., Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C., Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X., Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D., Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A., Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L., Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacble J.M., Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G., Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H., Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T., Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Swirskas R., Tector C., Turner R., Venter E.H., Wang A.H., Wang X., Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J., Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., Zheng X.-H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu S., Zhu S., Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
"The genome sequence of *Drosophila melanogaster*.";
Science 287:2185-2195(2000).
DR ENBL: AE0033442; AAF46302.1; -
DR FlyBase; FBgn0029985; CG1531.
DR InterPro; IPR001841; Znf_ring.
DR SMART; SM00184; RING; 1.
DR USQ SEQUENCE 1456 AA; 161358 MW; 61C55D04E4067893 CRC64;

	Query Match	4.6%	Score 420;	DB 5;	Length 1456;	
	Best Local Similarity	21.3%;	Pred. No. 2.6e-20;			
	Matches 210; Conservative	135;	Mismatches 342;	Indels	298;	Gaps 41;
QY	896	ERADTDNLNLTGMLQMAFHIALGILLKEEQOLQAPEEVTDFYH--KASRLGSSAMN	954			
Ddb	599	DSTMVDNCN---EASTSMA-----AAESTALAEGRSKRKNYHIWSSSRIGDGPFF	647			
QY	955	IOMLEKLKGIPQLGEGOKMITHLQMFDTVK--RL-----REKSLIVATTSGSESIKNDE	1009			
Ddb	648	IGNLLRKI-----AKQD--EQCAOSIDDIRARLPWNQRKQ----AEAKAREA-----	689			
QY	1010	ITHDKEAERKRKAEARHLRKHIMAQMSALOKNFETHKLMYDNTSEMPGREDSIMEEE	1069			
Ddb	690	----KEKEERRKK--AREQQRMQDMFANKQLFMQSAAASSSGMGVCPEDDEDDELVE	742			
QY	1070	STPAVSYSRIALGPKRGPSVTEKVLTCILCOEEQEVKIENNAVMLSACVKQSTALTQH	1129			
Ddb	743	EQPREKEYD-----CIICTTTP-STESNPICLVVLVE--SSGIUGH	781			
QY	1130	RGK-----PIELSGEALDPL-----FMDPDLAY--GYTT	1156			
Ddb	782	RRRTAERLPIPIINADESRLAHTRLAARENRTELLCLKFGDESWYLSNNWAYDNGVHV	841			
QY	1157	GSCGHVMHAVCWQKYFEAVOLSSQQRIHVDLFLESGEYLCPLCKSLCNTVIPIPL--Q	1214			
Ddb	842	QSCGGHHVHLSCLAYLKTLYTTORQPQV-----DRGEYFCPCVQLSNSVLPLSPQLDR	895			
QY	1215	PQKI-----NSENADAQAOLLTLARWTQTVLARIISGNIRHAKGENPIPIFFNQGMGSTL	1270			
Ddb	896	PTHLVRSNGPQPFERLVADLTDLIRENETI-----PPQTKUTEANGHAM--	938			
QY	1271	EFHSILSFVGESSIKYINSIKEMVILFATITVIRGLKVP-----PDER	1314			
Ddb	939	--EVMTNIAQRKVCKSSIIFRKLFIPTVTSIARTNLEAEIIQRGGSLCTANATRYKPKRE	995			
QY	1315	PRVPMWL-----TWSTC-----ATTIOAIENLLCDECKPLFG	1345			
Ddb	996	CIVPELLHVLVSVVVRVLVWPVLSWSASLAGLPVTATEPLPAHCHELIPSILADP-----	1049			
QY	1346	ALQNQRHGINKALMQFAVAORITPCQVLIQKHLVRLLSVVLNPNIKSEDPTCLLSIDLFHV	1405			

```

Db 1050 -----IALLLKEI-----LLAPLQLODDQFTFCWKV-----MTNLLYYQI 1084
Qy 1406 LVGAVLAPPSLYDDPDVLQPPSSVSSYNHLYLFHLITMAHMLQ----ILLTVDTGLPLAQ 1462
Db 1085 VVOLCVTLTDLCDHLVKVYGSTSVGSDN-----SAAESQQQESAAGCTTNNRRRAGQ 1136
Qy 1463 VQDESEEAHSASSFAETISQ-----YTSGSGICDIPGWLYLWVSLKNGIT--- 1506
Db 1137 QOQSSSLGKAMALV--LSQNTDLVHLRRDCTPSTSSSAAASAGSSSTTSTNHGASAA 1194
Qy 1507 -----PYLRCAALFFHYLLGVTPPELHTNSAEG--EYSALCS 1542
Db 1195 ASSATTIEVNLKSMELQALCLPFLURVALLRQHLYRHEMPE-----ISAPGLEFVRLVY 1250
Qy 1543 YLSLPTNLFLLFEQYWD-----TVRPLLRMC-----ADPALLNCLKQ 1580
Db 1251 YLELVTD-----SMDWDCFNASKGLCFIPGTETTLQFMCQQLMEVRPPADTVRELVLIN 1305
Qy 1581 KNTVVYRPRKRNLSIELPDDYSCLLNOASHFRCPRSADDERKHPVLCLFCGAILCSQNIC 1640
Db 1306 QHSLWQOQPR---LLELPREYERLEFY-YHERPCLNCYKVPKKESSICLLCGTIVCLKQNC 1360
Qy 1641 COEIVNGVEGACIFHALCHGAGVCFILKIRECRVVLVEGKARGCAYPAPLYDEYGETDP 1700
Db 1361 CAE-----NDCCBAVRHTLSCGGGIGIFLVVTSTYIIIVIRG-RACLWGSYLLDDFDEEDR 1415
Qy 1701 GLKRGNPLHLSRERYRKLHLVWQOH 1725
Db 1416 DLKRGKPLYLSDKRNLESQWLSH 1440

RESULT 13
Q9LZ94 PRELIMINARY; PRT; 795 AA.
AC 09LZ94;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE ECSRIFERUM3 (CER3).
GN T1E22.70.
OC Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I.; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA Bancroft I., Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL162874; CAB85536.1;
DR InterPro; IPR002114; PTS_HPR_ser.
DR InterPro; IPR001841; Znf_ring.
DR SMART; SM00184; RING; 1.
DR PROSITE; PS00589; PTS_HPR_SER; UNKNOWN_1.
SQ SEQUENCE 795 AA; 89971 MW; 401B2D08514B847D CRC64;

Query Match 3.7%; Score 343; DB 10; Length 795;
Best Local Similarity 23.0%; Pred. No. 2.5e-15;
Matches 190; Conservative 123; Mismatches 274; Indels 240; Gaps

Qy 1033 IMAQMS-----ALQKNFIETHKLMY-----DNTSEMPGKESIMEESTPAVSDYSRIA 1081
Db 56 LMLQLSDDTISESANWIESIKARLINCOTERSSDCRGCKDESME---SLEAMVQTVR 112
Qy 1082 LQPKRGPSVTEKVLTCILCOEEQVKEVIENNAMVLSACVOKSTALTQHRGKPIELSG--- 1138
Db 113 -----NKIENMINSLOTRVDHOPHEARN-----CSEKNSV-----GGPSTLOGRFP 153

```

```

QY 1139 -----EALDPLFMDPDLAYGTYTGSCGHVMAVCWQKYFEAVQLSSQORI 1183
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 154 DIRSQTSSRRPDAGSDGPHIDCD-----GVYSSCGHVAHOSCLERYLSKLSKRSGRRT 208
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1184 HVD---LFDLESGEVCLPCKSLCNTVTIPIIP-----LQPO-KINSENADALAOILLT 1231
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 209 VFEAGHIVDLKKFELCPVCRRLANSVLPCEPGDLCSVSKLQDSPRTKLRKDALOPSL- 267
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1232 LARWQTOTVIARISGYNIRHA-----KGENPIPIFFNOGMGDSLEFHSILSGVESS 1283
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 268 ---WLSEALCL-----LRSAAEVIEDGGRGKTVP-----QGDGPRR----- 301
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1284 IKYSNISKEMVILF-----ATTIYRIGLVKVPDPDRPRVPLMTWTSCAFTIOAIENLLG 1337
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 302 -KDLKSVSKMLWDFYFPKPEDTKLRWL--PPQS-----IVMWDTLKYSLSISME--IG 350
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1338 DEG-----KPLFGALONQHNGKLMQFAVAQR-----ITCQVOLI-QKHLVRL 1381
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 351 TRFAKNMPLVYCIDSLEYELKTSKGTLSVLLRVVQSSRTKNTIHRQRFVGMKHLAES 410
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1382 LSVLPLNKSIED-----TPCLLSIDL-----FHLVGAV--LAPPSLYW 1418
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 411 ICYGVSSSSSSSIFGSEGTGSLKNIDLLWNRASDPVLAHDPFSSLMWALFCLPFPFLTC 470
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1419 DDPVDLPQSSVSSYNHLYLFLHTMAHMLQILLTVDGLPLAQVQEDSEEAHSASSFFA 1478
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 471 EESL-----LSLVHIFHSVSLQVTIAYCACRP-----SELSELNFGENLLN 512
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1479 EISQYTSGS-----IGCDIPGWYLVWSLKNIGITPYLRCAALFFHLYLGVTPPE 1526
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 513 DISNALRESGWEYFRSNMNDLSCDIK-----TIRKYSLPFLRRCALLWKLL--KSTPR 565
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1527 ELHTNSAGEYSALCSYLSLPT--NLFLF--QEWYDVRPLQWCAADP--ALLNC-LK 1579
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 566 KLH-----EESDMFDLPSPDPTDNMDFIYSPQSELNHVQELKMFNIPPIIDILNDELL 619
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1580 QKNTVV-----RYPRKNS-----LIELPDDYSCLLNQASHFRCPSADDE 1620
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 620 RSSTQIWLQHFQREYRVNRVRSKLCITPVVFPQMLKPNLQDILLQRCIKRCV-NCTPR 678
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1621 RHPVLCLFCGAILCSQ--NTCCQEIYVNGEVEGACIFHALHCGAGVCIFLKIREFCRVVLV 1678
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 679 IEPPVCLLCSG--LCSPIWSPCCRE-----SGCPNHAIITCGAGTGVFLLR--TTILL 729
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1679 EGKARGCAYPAPYLDYGETDPLGKRNPLHLSRERYKRLHLVWQOH 1725
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 730 QRFARQSPWSPYLDTFGEEDIDMIRGKRLVNEERYAALTLYVGSH 776
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 14

```

ID Q96248 PRELIMINARY; PRT; 794 AA.
AC Q96248;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CER3 GENE.
GN CER3.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. WASSILEVSKAYA;
RX MEDLINE=96407843; PubMed=8811860;
RA Hannonfa A., Negruk V., Eisner G., Lemieux B.;
RT "The CER3 gene of Arabidopsis thaliana is expressed in leaves, stems,
RL roots, flowers and apical meristems.";
RL Plant J. 10:459-467(1996).
DR EMBL; X95962; CAA65198.1; -.

```

```

DR Mendel; 12959; Arath;2159;12959.
DR InterPro; IPR001841; Znf_ring.
DR InterPro; IPR002114; PHS_HPR_ser.
DR PROSITE; PS00589; PHS_HPR_SER; UNKNOWN_1.
DR SMART; SM00184; RING_1.
SQ SEQUENCE 794 AA; 89917 MW; CF5EB92AD42D8BAB CRC64;

```

Query Match

Best Local Similarity 3.6%; Score 335.5; DB 10; Length 794;
Matches 190; Conservative 123; Mismatches 273; Indels 241; Gaps 44;

```

QY 1033 IMAQMS-----ALQKNFIETHKLMY-----DNTSEMPGKEDSIMEESTPAVDYSRIA 1081
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 56 LMLQLSDDTISESANMIESIKARLIGNQTEKRSJGDKGKDESNME--SLEIAMYQTVR 112
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1082 LGPKRPSVTEKEVLCILCOEEQEVKIENNAWVLSACVOKSTALTQHRKPIELSG--- 1138
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 113 -----NKTIENINOSLTRVDHQHHEAN-----CSEKNSV-----GGPSTLOGRFP 153
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1139 -----EALDPLFMDPDLAYGTYTGSCGHVMAVCWQKYFEAVQLSSQORI 1183
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 154 DIRSQTSSRRPDAGSDGPHIDCD-----GVYSSCGHVAHOSCLERYLSKLSKRSGRRT 208
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1184 HVD---LFDLESGEVCLPCKSLCNTVTIPIIP-----LQPO-KINSENADALAOILLT 1231
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 209 VFEAGHIVDLKK-EFLCPVCRRLANSVLPCEPGDLCSVSKLQDSPRTKLRKDALOPSL- 266
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1232 LARWQTOTVIARISGYNIRHA-----KGENPIPIFFNOGMGDSLEFHSILSGVESS 1283
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 267 ---WLSEALCL-----LRSAAEVIEDGGRGKTVP-----QGDGPRR----- 300
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1284 IKYSNISKEMVILF-----ATTIYRIGLVKVPDPDRPRVPLMTWTSCAFTIOAIENLLG 1337
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 301 -KDLKSVSKMLWDFYFPKPEDTKLRWL--PPQS-----IVMWDTLKYSLSISME--IG 349
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1338 DEG-----KPLFGALONQHNGKLMQFAVAQR-----ITCQVOLI-QKHLVRL 1381
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 350 TRFAKNMPLVYCIDSLEYELKTSKGTLSVLLRVVQSSRTKNTIHRQRFVGMKHLAES 409
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1382 LSVLPLNKSIED-----TPCLLSIDL-----FHLVGAV--LAPPSLYW 1418
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 410 ICYGVSSSSSSSIFGSEGTGSLKNIDLLWNRASDPVLAHDPFSSLMWALFCLPFPFLTC 469
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1419 DDPVDLPQSSVSSYNHLYLFLHTMAHMLQILLTVDGLPLAQVQEDSEEAHSASSFFA 1478
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 470 EESL-----LSLVHIFHSVSLQVTIAYCACRP-----SELSELNFGENLLN 511
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1479 EISQYTSGS-----IGCDIPGWYLVWSLKNIGITPYLRCAALFFHLYLGVTPPE 1526
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 512 DISNALRESGWEYFRSNMNDLSCDIK-----TIRKYSLPFLRRCALLWKLL--KSTPR 564
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1527 ELHTNSAGEYSALCSYLSLPT--NLFLF--QEWYDVRPLQWCAADP--ALLNC-LK 1579
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 565 KLH-----EESDMFDLPSPDPTDNMDFIYSPQSELNHVQELKMFNIPPIIDILNDELL 618
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1580 QKNTVV-----RYPRKNS-----LIELPDDYSCLLNQASHFRCPSADDE 1620
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 619 RSSTQIWLQHFQREYRVNRVRSKLCITPVVFPQMLKPNLQDILLQRCIKRCV-NCTPR 677
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1621 RHPVLCLFCGAILCSQ--NICQEIYVNGEVEGACIFHALHCGAGVCIFLKIREFCRVVLV 1678
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 678 IEPPVCLLCSG--LCSPIWSPCCRE-----SGCPNHAIITCGAGTGVFLLR--TTILL 728
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1679 EGKARGCAYPAPYLDYGETDPLGKRNPLHLSRERYKRLHLVWQOH 1725
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 729 QRFARQSPWSPYLDTFGEEDIDMIRGKRLVNEERYAALTLYVGSH 775
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 15

```

Q9W3M5 PRELIMINARY; PRT; 783 AA.
ID Q9W3M5
AC Q9W3M5;

```


Result No.	Query	Score	Query			ID	Description
			Match	Length	DB		
1	1038	11.3	1958	1	UBR1_SCHPO	O60152 schizosacch	
2	914	9.9	2052	1	YDOB_SCHPO	O13731 schizosacch	
3	689.5	7.5	1950	1	UBR1_YEAST	P19812 saccharomyc	
4	647.5	7.0	1941	1	UBR1_KLULA	O60014 kluyveromyc	
5	208	2.3	2230	1	G0G4_HUMAN	K13439 homo sapien	
6	180.5	2.0	1829	1	MY5A_CHTCK	Q02440 gallus gall	
7	175.5	1.9	1427	1	REST_HUMAN	P30622 homo sapien	
8	175.5	1.9	1526	1	MY52_SCHPO	Q09i16 schizosacch	
9	175.5	1.9	2663	1	CENE_HUMAN	Q02224 homo sapien	
10	170	1.8	2298	1	C215_HUMAN	Q9y3r5 homo sapien	
11	168.5	1.8	3911	1	AKA9_HUMAN	Q99966 h a kinase	
12	164	1.8	2096	1	BP28_DROME	Q9vm75 drosophila	
13	163	1.8	1853	1	MY5A_MOUSE	Q99104 mus musculu	
14	158.5	1.7	1855	1	MY5A_HUMAN	Q9y4i1 homo sapien	
15	158	1.7	2033	1	EVPL_HUMAN	Q92817 homo sapien	
16	156.5	1.7	1790	1	USOL_YEAST	P25386 saccharomyc	
17	156.5	1.7	2710	1	TOXA_CLODI	P16154 clostridium	
18	156.5	1.7	2869	1	RBP1_PLAVB	Q00798 plasmodium	
19	155.5	1.7	1216	1	P1B1_RAT	P10687 rattus norv	
20	155.5	1.7	2704	1	BPA1_HUMAN	Q03001 homo sapien	
21	154	1.7	3685	1	DMD_HUMAN	P11532 homo sapien	
22	153.5	1.7	1875	1	MLP1_YEAST	Q02455 saccharomyc	
23	153	1.7	1682	1	MSP1_PLAF3	P19598 plasmodium	
24	152.5	1.7	1325	1	GL60_MOUSE	P59537 mus musculu	
25	152.5	1.7	3433	1	UTRO_HUMAN	P46939 homo sapien	
26	151	1.6	3660	1	DMD_CHICK	P11533 gallus gall	
27	151	1.6	5430	1	ACF7_HUMAN	Q9upn3. homo sapien	
28	150.5	1.6	1539	1	V373_HUMAN	O15078 homo sapien	
29	150.5	1.6	2469	1	TEGU_HSVYA	Q01056 herpesvirus	
30	150.5	1.6	5327	1	ACF7_MOUSE	Q9qxz0 mus musculu	
31	149.5	1.6	1453	1	V373_BOVIN	Q9tu23 bos taurus	
32	149.5	1.6	1679	1	YF09_YEAST	P40457 saccharomyc	
33	149.5	1.6	2017	1	MY5N_DROME	Q99323 drosophila	

QY 100 GRVKSGETTYSRCDCAIDPTCVLCMDCFOQSVHKNHRYKMHTSTG-GFCDGCDTEAWK 158
Db 97 GHIFKGEVFRCKTCSVDNSALCVKFRATSHKHDTSTFVSAGSGCCDCGNAAWI 156
QY 159 TCFPCVNEPGRAGTIKENSRCPLNEEVIQARKIFPVSIVKIVVEMTIWEER--KELPPE 216
Db 157 GDVSKTHSHEEDATISNDMIDEIPEKLENSIQTTIDCVLDFVLDVFCSPENLKKMPTL 216
QY 217 LOI--REKNER-----YYCVLFNDEHHYSVHVIYSQRLACELAE---A 256
Db 217 ESILQDEKTSRLSENKYGDIDSCNMYSVLVWLNDEKHSFKQFQIITLAL--ELPNVFG 274
QY 257 QLHTTAIDKEGRRAKAYACQAEKEDIK---SHSENVSOHPLHVEVLHSEIMAHQKF 313
Db 275 KKMNIINDIGR-----ACIVETNIKELKIGOKLAQINLAVSIRSMRDFREES 325
QY 314 ALRGLSMNKMITSYS-----SDFRQIFCOACLR----- 341
Db 326 CAVLEWLADIAGSICCKRNYFSSVICKELVRPWCNGLHNSDLTFRLSRLALPEIVA 385
QY 342 -EEDP-----SENCLISRLM----- 356
Db 386 IDSPDIFLNEDHINSGSPDSTSHMLTDESSIHSHRWYPSNSLPDVLASYASRVDFYFF 445
QY 357 LWDAKLYKGARKILHELIFSFFMEMEKKLIFAMEFYKYLQLOKEYISDDHRSISITA 416
Db 446 LYDLKMLSLRYKLOELVGYFITQPGKEIMGARIAISYRLAELFLLDDREPHSVIF 505
QY 417 LSVQMTFTPLARHLIEBQNVISVITETLLEVL-----PEYLD-----RNNKENFGQY 464
Db 506 FPMQIFTVADVAKLLVTEYDFLTINATLYFTFYKLNTPYVDQHAMIRTDSAF-- 563
QY 465 SODKLGRIYAVICDLKYILISKPTIWTREL--MOLFEGFRSFLKILTCMOGMEIRROV 522
Db 564 -----SRRYIHIFHHIQMSPICPV-AEIVREDLKFLOAYADFFNL---FGCMCPYTRAV 614
QY 523 QOHEVDPP-WEAATAIQMOL-----KNILLMPQEW-----CACDEELLVAYKECHRAV 571
Db 615 SCHWENDSMYVLNLSQVAKCRHGVNFMELNTKLANAINYLILPYKARNES 674
QY 572 MRCSTSFSTSSKTVVOSGCHLSKTSYRVEDLSIHLPLSRTLGLHVLRLGAVSRL 631
Db 675 W-TWNETSGTITVDER--GNS-KLIEYDIALQPVSFHPLHLWLVL-----L 719
QY 632 HEFVSFEDF-----QVEYL--VEYPLRCLVLVAQVVAEMRRNGLSLSQVFIYQDVCKR 684
Db 720 SFYVERONYKLLWTOLDLAVTDHPLRYCAMLQSMRKLWIRNGTTLRDQAHYHNLSPH 779
QY 685 EEMYDKIIMIQIGASLMDPNKFLLLVQRYELAEAF--NKTISTKQODLIKQYNTLIEE 742
Db 780 EYTFDLVLLQLTLTYGDPDAILPSFSRFOLEDQMGYGFVPHKHVD-VSQVTIMBEE 838
QY 743 MLOVLIYIGERYPGVGNVKEEVTMEITHLICIEPMPHSAIAKNLPENNETGLEN 802
Db 839 FLLLLISIVCNVAVLDHWDITRR--IEYGAHILCFRPLPYSEITKRCHELLEHKQFES 896
QY 803 VINKVATEKK--PGVSGHGVYELKDESLDFNNFYHYSKTOHQSAHEMKKR--RKQENKD 860
Db 897 TLKQVATERNAEGINDSGSFTLKDEYFDVDPFNHLYSRNQREAEANILRRYSKQSHK 956
QY 861 EALPPPPPPPCPAFSK---VINLNCIDIMVIL--RTVFERAIDTDSNLWTEGMLQMAFHI 917
Db 957 --LESVVYEEYHPILHNSITIPILQSDSFVGLWHTIYVAYIYPYDQGLKGLVNTALHA 1014
QY 918 LALGLLEKQOLQKAPEEVTFDYHKKASRLGSSAMNTOMLLEKLGIPQLEGQD---- 973
Db 1015 CLLVLMSEK-----GSEPIFSKKICE--NRFPVVEGLEQYCN 1050
QY 974 ---MITWLOM-----FDTVKRLREKSLIV-----ATTSGSESIKND 1009
Db 1051 PDVTLFSVLCOMKHNURNVYVK---EKISLIMKILKSEVPLLYEPVYAEIUSISSKIV 1107
QY 1010 ITHOKEAERK--RKAEEARLHROKIMAOALSALQKNFIETHKLMYDNTSEMPGKEDSIM 1068

RESULT 2

YDOB_SCHPO
ID YDOB_SCHPO STANDARD: PRT: 2052 AA.
AC O13731;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL 234.0 KDA PROTEIN C15A10.11 IN CHROMOSOME 1.
GN SPAC15A10.11

Qy	504	SFLKILTCMOGMEETIRROVGOHI--EVPDWEAAIAIQMLKNILLMPQWC-----ACDE	557
Dd	657	SLLTCLCKFNGAWKIKRGEGBHVLHQDNFTSYLEYTTSYIIQTAEKVSEKSDSIDS	716
Qy	558	ELLIVAYKECHKAVMRCSTSISSSKTVQSCGHSLETKSVRSVEDLVSHLPLSRFLAG	617
Dd	717	KLFNA-----IRIITSSFLGNRSUTYKLIDYSHEVIKPSVSHERVAFMNPLQWLSF	768
Qy	618	LHVRLSRLGASRLH---EFVSFEFOVEVLVEYPLRCLVLVAQOAWEMRRNGLSLISO	674
Dd	769	LIEKVSJKDAYEALEDCSDFLKISDFS-----LRSVLCSDOIDGVFWRWGMVSLHQ	820
Qy	675	VFIYDVOKCREM---YDKDIIMLQIGA--SLMDPNKFLLLVLQRYELAEAFNKTIISTKDQ	730
Dd	821	ASYV---KNNPELGYSRDIHLNOLAILWERDDIPRIIYNILDRWELLDWFTGEVDYQHT	877
Qy	731	DLIKQNTWLIEMLQVLIYIGERYPCVGWNTE- -EVTMRIIHLLCIEPMPHSAIA	787
Dd	878	VYEDKITSFIIOQIFAIYOILITERQYFKTFSLKDRMDQIKNSIINYLMKPLSYSKUL	937
Qy	788	KNLPEN-ENNETGLENIKNVATKKP-GVSGCHGVYELKDESLEKDFNMVFHYVSKTOHKS	845
Dd	938	RSPVDLYIEDTFEDEALEEVSFVEPKGLADNVGFKL-ASL-----YAKVDPCLK	987
Qy	846	AEHMQK-----KRKQKNDKEALPPPPPPEFCFPASKVINLLNCDIM---MYILRT	893
Dd	988	LLNELENPESSATIISKHLAKDKDEIAKVVLIPOVS-----IKQLDKDALNLGAFTNT	1041
Qy	894	VFERAI-----DTDSNLWTEGMLQMAFHILALG-LLEBKQOLQA-----PEEEVTFD	940
Dd	1042	VFAKYVKILLQVCLODMEDSTFLNELHLHVGHIFRDDELINGKDSIPAEYLSKPTICNLLS	1101
Qy	941	FYKASRLGSSAM--NIQMELLEKLGPOLBEGQXDMITWILOWFDTVKRLREKSLIVAT	998
Dd	1102	IANAKSDFPSSIVRKADYLEKKM-----INMKNELFESLIA--	1139
Qy	999	TSGSEIKNDEITHDKK-----AERKRAEAAARLHRQIMAQMSALQKNFTETH	1048
Dd	1140	----SFGNOYVNDYKDKKLQGVNLQETERKERRRLAKKHQARLLAKFNNOQT KPMKEH	1194
Qy	1049	KLMYDNTSEMPCKEDSIMBESTPAVSYSRIALGPRGPSVTEKEVLTCLCOEQEOEVK	1108
Dd	1195	ESSFD--EQDNDVDMV-----GEKVYESEDFTCALQDSS---	1227
Qy	1109	IENNAMVLSACVOKSTALTQHRGKPIELSGEALDP-----LFMDPDLA	1151
Dd	1228	-STDFFVIPA-----YHDHSPFRPGNIFNPENPMWDGFYNDDEKQAYIDDVVL	1277
Qy	1152	YG-TYTGSOG-----HYMHAVCMOKPYEAVQLSSQQRIHVDLFDJESGYLCPCLK	1201
Dd	1278	EALKENGCGSKRPVPSCNHHIHNCFKRYVQKKRFS-----SNAFICPLCQ	1324
Qy	1202	SLCNTVIPITPIOPQOKINSENADALAQLLTARWIOQTVLARISYINTRHAKGENPTDIFF	1261
Dd	1325	TFSNCTPLP--COTSKANTG-----LSLDMFLESLSL-----	1355
Qy	1262	NOCMGDBSTLEFHSILLSFCGVSSIKEYSNSIKEMWTLFATTIYRIGLKVPPDEROPRVMLT	1321
Dd	1356	-----DTLS--RLPKPTEENTYNTINSIFSAMI-----	1381
Qy	1322	WSTCAFTIAIENLLGDGCKPLFGALQNQRBNGLKALMQFAVAORITCPOVLI-----Q	1375
Dd	1382	-SOCQGFDAKVRKANRFSSHKDVLSILSVHWANTISML--EIASRLEKPYSISPFERSREQ	1437
Qy	1376	KHLVRLLSVVLNPKTESDTPCLLSIDLPHVLUGVAVLAPPSLYWDDPDVDLOPSSVSSSYNH	1435
Dd	1438	KY-----KTLKNI-----LVCIMLFTFYIGR---PSMEF-EPYQPQPDVTWNO-NQ	1478
Qy	1436	LY-----LEHLTMAHMLQILLTVDTGPLAQAOVEDSEAHSSAFFAEISQ-----	1482
Dd	1479	LFQYIVRSALFPSVLSRGTVTEALTFTSRQPLRDFLOGLSDAEQVTKLYAKASKIGDVLK	1538

Qy	1483	-----YTSGSIGCDIPGWVLVSLKNGITPYLRCAALF---FIWLL	1520
Dd	1529	VSEQLFAALTISDVMEGLDSEIIYDIAYTFLLKSL-----LPTIRRCILVFIKLVHEL	1594
Qy	1521	GVTPEELHTNSAEGEYSALCSYLSLPNLFLLFOEWDTVRPQLQRWCADPALINCLKQ	1580
Dd	1595	KDSENETLVINGHEVEEE-----LEFDATAEFNKALKMITEKESLDVLDTT	1641
Qy	1581	KNTVRYRPRKRN-----SLIELPDDYSCLLNQASHFRCPRSAD-----DERKHVP	1625
Dd	1642	QESIIVSHPYLENIPYECGIIKLIDLSKLYNTVYTOSKEIKLBRESRQHMKNDNRIDFK	1701
Qy	1626	LCLFCGAIL-----CSQNICCOEIYNGBEVGACIFHALHCGAGVCIFLK	1669
Dd	1702	ICLTCGVVHLRADRHMTKHLKNCKFKPGAFMPNSE-----VCLHLT	1747
Qy	1670	IRECRVIVEKGARGCAYPAPYLDEYGETD-PGLKGNPLHLSRRYRKLUHLVWQQHCII	1728
Dd	1748	OPPSNIFI-----SAPYLSHGCEVGRNAMRGDLTTNLKRYEHLNRLWINNEIP	1797
Qy	1729	EETARSQE-----TNQMLEGFN	1745
Dd	1798	GYISRVMGDFFRVTLNSGLFLFAFN	1822
<hr/>			
RESULT 4			
ID	UBRL_KJULA	STANDARD;	PRT; 1941 AA.
AC	O60014;		
DT	15-DEC-1998 (Rel. 37, Created)		
DT	15-DEC-1998 (Rel. 37, Last sequence update)		
DT	15-DEC-1998 (Rel. 37, Last annotation update)		
DE	N-END-RECOGNIZING PROTEIN (UBIQUITIN-PROTEIN LIGASE E3 COMPONENT) (N-RECOGNIN).		
DN	UBRI.		
GC	Kluyveromyces lactis (Yeast).		
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
OC	Saccharomycetales; Saccharomycetaceae; Kluyveromyces.		
OX	NCBI_TaxID=28985;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Waller P.R.H., Varshavsky A.;		
RL	Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.		
CC	-!- FUNCTION: RECOGNITION COMPONENT OF THE N-END RULE PATHWAY. BINDS TO PROTEINS BEARING AMINO-TERMINAL RESIDUES THAT ARE DESTABILIZING ACCORDING TO THE N-END RULE, BUT DOES NOT BIND TO OTHERWISE IDENTICAL PROTEINS BEARING STABILIZING AMINO-TERMINAL RESIDUES.		
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on use by non-profit institutions as long as its content is in no modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcements or send an email to license@isb-sib.ch).		
CC	EMBL; AF061534; AAC15841.1; -		
DR	InterPro; IPR003126; znf.UBR1.		
DR	InterPro; IPR001841; znf.ring.		
DR	Pfam; PF02207; zf-UBR1; 1.		
DR	SMART; SM00184; RING; 1.		
DR	SMART; SM00396; Znf-UBR1; 1.		
DR	Ligase; Ubiquitin conjugation.		
SK	SEQUENCE 1941 AA; 223682 MW;		
SQ	37C7C2E1BCAA0803268 CRC64;		

Query Match 7.0% Score 647.5; DB 1; Length 1941;
Best Local Similarity 20.5%; Pred. No. 7.6e-32;
Matches 409; Conservative 319; Mismatches 689; Indels 581; Gaps 91

Db 1397 SSLRKQYDEEKCELLQVQDLQVDFKVDLSKEKISALEQVDDWS 1439

```

RESULT 6
MYSA_CHICK
ID MY5A_CHICK STANDARD; PRT; 1829 AA.
OC Q02440.
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MYOSIN VA (MYOSIN 5A) (DILUTE MYOSIN HEAVY CHAIN, NON-MUSCLE) (MYOSIN
GN MYOSA.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=93012002; PubMed=1383040;
RA Sanders G., Lichte B., Meyer H.E., Kilimann M.W.;
RT "cDNA encoding the chicken ortholog of the mouse dilute gene product.
RT Sequence comparison reveals a myosin I subfamily with conserved C-
RT terminal domains.";
RL FEBS Lett. 311:295-298(1992).
RN [2]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1154-1163.
RC TISSUE=Brain;
RX MEDLINE=93107155; PubMed=1469047;
RA Espreafico E.M., Cheney R.E., Matteoli M., Nascimento A.A.,
RA De Camilli P.V., Larson R.E., Mooseker M.S.;
RT "Primary structure and cellular localization of chicken brain
RT myosin-V (p190), an unconventional myosin with calmodulin light
RT chains.";
RL J. Cell Biol. 119:1541-1557(1992).
CC -!- FUNCTION: PROCESSIVE ACTIN-BASED MOTOR THAT CAN MOVE IN LARGE
CC STEPS APPROXIMATING THE 36-NM PSEUDO-REPEAT OF THE ACTIN FILAMENT.
CC MAY BE INVOLVED IN MELANOSOME TRANSPORT, OR ALTERNATIVELY, IT MAY
CC BE REQUIRED FOR SOME POLARIZATION PROCESS INVOLVED IN DENDRITE
CC FORMATION (BY SIMILARITY).
CC -!- SUBUNIT: MAY BE A HOMODIMER, WHICH ASSOCIATES WITH MULTIPLE
CC CALMODULIN OR MYOSIN LIGHT CHAINS.
CC -!- SUBCELLULAR LOCATION: GOLGI-DERIVED CYTOPLASMIC MEMBRANES
CC (POTENTIAL).
CC -!- TISSUE SPECIFICITY: NEURONAL AND NONNEURONAL CELLS OF THE BRAIN.
CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -!- SIMILARITY: CONTAINS 6 IQ DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 DILUTE DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; X67251; CAA47673.1; -.
CC EMBL; Z11718; CAA77782.1; -.
CC HSSP; P08799; 1MND.
CC InterPro; IPR002710; DIL.
CC InterPro; IPR000048; IQ.
CC InterPro; IPR001609; myosin_head.
CC Pfam; PF01843; DIL; 1.
CC Pfam; PF00612; IQ; 6.
CC Pfam; PF00063; myosin_head; 1.
CC PRINTS; PR00193; MYOSINHEAVY.
CC ProDom; PD000355; myosin_head; 1.
CC ProDom; PD003376; DIL; 1.
CC SMART; SM00015; IQ; 6.
CC SMART; SM00242; MYSC; 1.

```

```

DR PROSITE; PS50096; IQ; 6.
KW Myosin; Repeat; ATP-binding; Calmodulin-binding; Actin-binding;
KW Coiled coil; Phosphorylation.
FT DOMAIN 1 766 MYOSIN HEAD-LIKE.
FT DOMAIN 767 789 IQ 1.
FT DOMAIN 790 814 IQ 2.
FT DOMAIN 815 837 IQ 3.
FT DOMAIN 838 862 IQ 4.
FT DOMAIN 863 887 IQ 5.
FT DOMAIN 888 915 IQ 6.
FT DOMAIN 916 1239 COILED COIL (POTENTIAL).
FT DOMAIN 1315 1419 COILED COIL (POTENTIAL).
FT DOMAIN 1661 1766 DILUTE.
FT NP_BIND 163 170 ATP (BY SIMILARITY).
FT DOMAIN 644 666 ACTIN-BINDING (POTENTIAL).
FT MOD_RES 1734 1734 PHOSPHORYLATION (POTENTIAL).
FT CONFLICT 1142 1142 E -> EQ (IN REF. 2).
SQ SEQUENCE 1829 AA; 212381 MW; 0538B278DFC09F6E CRC64;

Query Match 2.0%; Score 180.5; DB 1; Length 1829;
Best Local Similarity 17.7%; Pred. No. 0.0054;
Matches 312; Conservative 269; Mismatches 633; Indels 545; Gaps 85;

QY 94 GAFOLCGRVFKSGTYSYCRD---CAI-----DPTCVLC-----MDCFQDSVHKHNRKMHIT 142
DB 324 GIFRILAGILHLGNVEFASRSDSCAIPPKHDPITFCDLMGVD-YEEMAHWLCHRKLAT 382
QY 143 STGGFCDCGDEAWKTGPFVFNHPEGRAGTIKENSRCPLNEEVIVQARKIPSPVTK--- 199
DB 383 A-----TEYI-KPISKLHAINARDALAKHIYANLFWIVDHVKNKALHSIVKQHS 431
QY 200 YVEMTIWEKEKLPPELQIREKNERYCVLFNDE--HHSYDHYIVYLSQALDCELAEEA 257
DB 432 FIGVLDIYGPE-----TPEINSFEQ--PCINYANEKLOQQFNHMFVKLEQE---EYMRQ 481
QY 258 LHITTAIDKEGR-----AVKAGAYAAQCEAKEDIKSENVSQHPHVEVLHSEIMAHQ 311
DB 482 IPWTLDFYDQPCINLEAKMGVLDLDECKMPKSGDDTWAQKLYNTHL-----N 533
QY 312 KFAL-----RLGSMWNK--IMSYSDFRQIFQACILREPDSENPCLISRLMLWDKLYKG 365
DB 534 KALFEKPRLS--NKAFIIKHFADKVEYQCEGFLEKKNDT-----VVEE 575
QY 366 ARKILHELIFSSPFMEYKKLFAMEFYVYKQLOKEVISDDHDSISITALS-----VOM 421
DB 576 QIKVL-----KSSKKFKLLPE--LFQDEEKAIPTSATPSGRVPL 613
QY 422 FTVP-----TLARHL-----IEQNVISVITETLEVLPEYLD--RNKNFQGY 465
DB 614 SRTPVKPAKARPGGTSKEHKKTGVGHQFRNSLHLLMETLNATTPHYVRCIKPNDKPEPTF 673
QY 466 QDKLG-----RVYAVICDLKYILISKPTIWERLMOFLGFRSFLK---IL-----TCMQ 513
DB 674 DEKRAVQOLRACGVLETIRISAAGFPSPRWYQ---EFFSRVRLMKQKDVLSDRKQTCN 730
QY 514 GMEIR-----RRQVG-----OHIEVDPDWEAAIAIQMLKNILLMFQEWCA 554
DB 731 VLEKILDKDKYQFGKTKIFFRAGOVAYLEKIRADKLRAACIRIQKTIRGWLMRKKYMRM 790
QY 555 CDELLLVAYKECHKAVMRCSTSFISSKT--VQVSCGHS-LETKSYRVSEDLVSIHPL 611
DB 791 RRAAITIQRYVVRGHQA--RCYATEFLRRTRAIIIIQKFRVMYVVRKRYQCMRDAT---IAL 845
QY 612 SRTLGLHVLRLSRLGAVSRHLEFVSFEDQVEVLVEYFLRCLVLVAOVVAEWRNGLSL 671
DB 846 QALLRGYLVLR-NKYOMMLREHKSI-----IIQKHVRCW----- 877
QY 672 ISOVFYQDVKVC-----REEMYDKDIIMLOIGASLMDPNKFLLLVLQRYELAEAFNK 723
DB 878 LARVHYHRTLKAIYVYLQCCYRRMMAKRELKCLKTEARSVERYKKLHGLE-----NK 929
QY 724 --TISTKDQDLIKOYNLTIEEM--LQVLIYIVGERYVPGV-----GNVTKEEVTM 769

```

Db 930 IMQQRIDEQNKYKSLLEKMMNLEITYTETETKLRSDVERLRMSDEEAKNATNRVLSL 989
Qy 770 REIHLICIEPMPSALAKNLPENNETGLEINVAVTFKPGVSGHGVYELKDESL- 828
Db 990 QEEI-----AKLRKHOTOTEKKYIEWADYKHETEOLVS-----ELKEONTL 1034
Qy 829 -----KDFNMVYHYKSTQSHAKHMQK---RRKQ---ENKDRALPPPPPPPECFPAFSK 877
Db 1035 LKTEKEELNRRIHQAK---EITETMEKKLVEETKQLELDLNDERL-----R 1078
Qy 878 VINLLN-----CDIMYILRTVFERAIDTDSNLTWTEGMLQMAFHILA--- 919
Db 1079 YONLLNFEFSLEERYDDLKDEMNLVSPKPGHKRTDTHSS--NESEYTFSSITEAED 1136
Qy 920 LGLEEKQOQLOKAPPEEVTFDFVHKASRLGSSAMNIOMLEKLGIPQLEGOKDMITWIL 979
Db 1137 LPLRMEPESEKAP-----LDMSLFKLQKRVTELEOEKOSLO--- 1174
Qy 980 QMFDTVARLREKSLIVATTSGSESIKNDIETHDKAEAKRRKRAEAARLHRQKIMQMSA 1039
Db 1175 --DELDRKEQALRAKAKEERPPIRGALEYESLK---ROELSES---ENKLLKNELE 1225
Qy 1040 LQKNFIETHKLYDNTSEMPGKEDSINEEESTPAVSYSRIALGPKRGSPVTEKVLTCI 1099
Db 1226 LQKALTETR-----APEVTAPGAPAY-RVLL-----DQUTSV 1256
Qy 1100 LCQEQEVKLENNAMVLSACVQKSTALTOHRGKPIELSGALDPLFMDPDLAYGTYTGSC 1159
Db 1257 --SEELEKKEEYILRLSRLVSQKEAF-----QPKDKNTMTSTILLEVDQKMKDGEI 1309
Qy 1160 GHVMHVMCKYFEAVQSLSSQRIHVDLFLESGEYLCPLCKSL-----CNTVPIPI 1212
Db 1310 AQAVIGLKTNRLLLESQKQKSHENELESRLGE-----IQSLKEENRQOQLAONLQ 1364
Qy 1213 LQPO-----KINSNADALAQL-----LTLARWITQTVLARIKSGYNIRHAKG 1253
Db 1365 LPPEARIEASLOHEITRLTNELDLMEQLEKQDKTVRKLLKQKLVFAKKIGLEVQGMEN 1424
Qy 1254 ENPIPFNGMGDSTL-----EFHSILSFGVSESIK----- 1285
Db 1425 ISPGQII-----DEPIRPVNIPIRKEKDFQGMLEYKKEDEQKLVNILELKPGRVAVNL 1478
Qy 1286 -----YSNSIKEMVILFATTVIRIGLKVPPDERPRVPMLTW--STCA 1326
Db 1479 IPGLPAVILEMCMVRHADYLNDDQKVRSLTSTINGI-KKVLKRGDDFETVSVFWSNTR 1537
Qy 1327 FTTOAIBENLGDGKPLFGALQNRHNGKALMQFAVA---QRITCPQVLIQRIHLVRLS 1383
Db 1538 F-LHCLAKQSGEGGFMKHNTPRQNEH---CLTNFDLAERYQVLSDLAIQIYQOLVR--- 1589
Qy 1384 VLPNKSIEDTPCLLSIDLPHVLGAVLAPSLYWDVDPVQLPSSVSSSYNHLFLHLIT 1443
Db 1590 -VLENI-----LQPMIVSCMLEH----- 1606
Qy 1444 MAHMLQILLTV-DTGLPLAQVOEDSEEAHSASFFAISIYTSIGSICDIPGWVLYSLK 1502
Db 1607 --ETIQGVSKVFTGL-----RKRTSSIADEGYTTLDSITIRQNSRHS-VMCQ 1651
Qy 1503 NGITPYL-RCAALFFHYLLGVTPPEELHTNSAEGEYSALCSY---LSLPTNLFLLFOEYW 1558
Db 1652 HGMDPELIKQVVKQMYIIGAVTLNLLLR-----KDMCSWSKGMQIRYNV----- 1697
Qy 1559 DTVRPLLQRCADPALLNCLKQNTKQNTVVYPRKRNLSIELPDDYSCLLNQASHFCRPSAD 1618
Db 1698 -----SQLEEWLRDKNLMNS-GAKETL-----EPLIQAAQLLQVKKKTD 1735
Qy 1619 DERKHPVCLFCGAILCSQ 1637
Db 1736 EDAAE-AICSMCNALTAQ 1752

RESULT 7

REST_HUMAN
ID REST_HUMAN STANDARD; PRT: 1427 AA.
AC P30622;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 20-APR-2001 (Rel. 40, Last annotation update)
DE RESTIN (CYTOPLASMIC LINKER PROTEIN-170 ALPHA-2) (CLIP-170) (REED-
DE STERNBERG INTERMEDIATE FILAMENT ASSOCIATED PROTEIN).
OS RSN.
GN Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Peripheral blood monocytes;
RA MEDLINE=92289675; PubMed=1600942;
RX Bilbe G., Delabie J., Brueggem J., Richener H., Asselbergs F.A.M.,
RA Cerletti N., Sorg C., Odink K., Tarcsay L., Wiesendanger W.,
RA de Wolf-peeters C., Shipman R.;
RT "Restin: a novel intermediate filament-associated protein highly
RT expressed in the Reed-Sternberg cells of Hodgkin's disease.";
RL EMBO J. 11:2103-2113(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC MEDLINE=92405160; PubMed=1356075;
RX Pierre P., Scheel J., Rickard J.E., Kreis T.E.;
RA "CLIP-170 links endocytic vesicles to microtubules.";
RT Cell 70:887-900(1992).
CC -!- FUNCTION: SEEMS TO BE A INTERMEDIATE FILAMENT ASSOCIATED PROTEIN
CC THAT LINKS ENDOCYTIC VESICLES TO MICROTUBULES.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC, ASSOCIATED WITH THE
CC CYTOSKELETON.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THE REED-STERNBERG CELLS
CC OF HODGKIN'S DISEASE.
CC -!- SIMILARITY: CONTAINS 2 CAP-GLY DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X64838; CAA46050.1; -;
CC EMBL; M97501; AAA35693.1; -;
CC PIR; S22695; S22695.
CC MIM; 179838;
CC InterPro; IPR000938; CAP-Gly.
CC InterPro; IPR001878; Znf_CCHC.
CC Pfam; PF01302; CAP_GLY; 2.
CC SMART; SM00343; ZnF_C2HC; 1.
CC PROSITE; PS00845; CAP_GLY_1; 2.
CC Cytoskeleton; Microtubules; Coiled coil; Alternative splicing.
KW DOMAIN 60 125
FT DOMAIN 143 204
FT SER-RICH.
FT DOMAIN 214 279
FT CAP-GLY 2.
FT DOMAIN 304 331
FT SER-RICH.
FT DOMAIN 350 1342
FT COILED COIL (POTENTIAL).
FT DOMAIN 1408 1421
FT CCHC-BOX.
FT VARSPIC 457 491
FT MISSING (IN SHORT ISOFORM).
FT CONFLICT 1069 1069
FT D -> E (IN REF. 2).
SQ SEQUENCE 1427 AA; 160989 MW; 0A4F166DD94254E8 CRC64;

Query Match 1.9%; Score 175.5; DB 1; Length 1427;
Best Local Similarity 19.4%; Pred. No. 0.0077;
Matches 200; Conservative 160; Mismatches 344; Indels 329; Gaps 45;
QY 167 EPGRAGTIKENSRCPLNEEVIVQARKIPPSVIKYVVVETIWEELPPELO-----IREK 222

```
Db 488 EDTRVATVSEKSR-----IMLEKDLALRVOEVAELRRR 521
QY 223 NERYCVLFENDEHSHYDHYIYLSQALDC-----ELAAQLHTTAIDKQGRRAVAKAGA 275
Db 522 LESNKPAGDVDMSSLQLEISSIQEKLEVTTRDQREITSLSKEHFGAREETHQKEKA-L 580
QY 276 YAACO---EAKEDIKS-----HSENVSOHLPHVEVLHSEIMAHQFAURLG-SWNKNIMS 326
Db 581 YTATEKLSKENSLSKLEHANKENSALTVAIKWKSLETAIAHQOAMEELKYVFSKGLGT 640
QY 327 YSSDPRQJFCO-----ACLRPEPDSENCLISRLMLWDKAKYKGARKIILHELI 374
Db 641 ETAEPALKTQLEKRWLDYOHIEINLQOODSER----- 674
QY 375 FSSFFMEMEYKFLAMEFYKQKQKQKEYISDDHRSISITALSVMETVPTPLARHLIEE 434
Db 675 -AAHAKEMEARAKLMKYIK-EKENSLEAIRSKLOKAE-----OHLVEM 717
QY 435 QNVISVITET-----LLEVLPYLDNNKF--NFQGY---SODKGRVYAVTCDLKYILI 484
Db 718 EDTLNKLOEAEIKVELEVLQAKCNEQTKVIDNFTSQLKATEEKL-----LDL----- 765
QY 485 SKPTIWTBRLMQLFEGPSFLKILTCMOGMEIIRQVCOHLEVDPDMEA--AIALQML 542
Db 766 -----DALRKASSEG-KSEMKKL--ROOLEAABEKQI-KHLEIKNAESSKASSITREL 814
QY 543 KNILLMFQWACDBELLIVAYKECHKAVMRCSTFISSSKTVWQSCGSHLSFTKSVRYSE 602
Db 815 QG-----RELKLTNLOE-----NLSEVSQVKETLEKLOLKEFAEASE 854
QY 603 DLVSTHLPRLSRTLAGHVLRLSRLGAVSRHLEFVSPEDFQVEVLVEYPLRCLVLVAQVVAE 662
Db 855 EAVSVQSRMQETVNLHQEEQFNMLSS-----DLEKL----- 887
QY 663 MWRRNGLSLSIOVFYOD-----VKCREMYDKDI--IMLOIGASLMDPNKFLLLVLQR 714
Db 888 ---RENLDAMEAKFREKDRBEOQLKAKEKL-ENDIAELMKNGSONSSOLT----- 935
QY 715 YELAEAFNKTISTKQDDILKQYNTLIEEMQLVLIYVGRYVPGVNTVKEVMTWREIHH 774
Db 936 -----MDELRLKERDV-----BELQLKLTKANEN--ASFLQKSIEDMTVK----- 974
QY 775 LCIEMPHPHSAITAKNLPENNETGLENVINKVATFKPGVSGHGYELKDESLEKDFNNY 834
Db 975 ---ABQSOQEAAKHKEEKKELERKLSDEKKMET-----SHNQOQELKAR----- 1017
QY 835 FHYSKTQHSKAEHMQKRRKROENKDEALPPPPPEFCPAPSKVINLNCNDIMMYILRTV 894
Db 1018 --YERATSETKTKEE-----ILQNL 1036
QY 895 FERAIDTSNLWTECMLQMAFHILALGLLEEKQOQLOKAPEEVTFDYHKASRLGSSAMN 954
Db 1037 QKTLTDTEKL--KGAREE-----NSGLLOLELELRKQADKA-----KAAQTAEADAMQ 1082
QY 955 I--QMLLEKLGIPOLEGOKDMITWLOFDTVVKRLRKESCLIVATTSGSESIKNDEITH 1012
Db 1083 IMEQWTEKETETLASIEDTKQTNAKLQNLDELDTLKENNLKNVEELNKSRELLTVENQKWE 1142
QY 1013 DKEAERKKAFAARLHRQK IMAQMSALQKNFTIETHKLMYDNTSEMPGKEDSI-----ME 1067
Db 1143 FRKETETLKQAAQK-----SQQLSALQE---ENVKL-----AELGRSRDEVTSHQKLE 1189
QY 1068 EESTAVSDYSRIALGPKRGPSPVTEKEVLTCILCQEEQEVKTENNAMVLSACV-QKSTAL 1126
Db 1190 EER-----SVLNNQLLEMKKRESKFIDK-----DEEKALQKSISITALLTEKDAEL 1238
QY 1127 TOHRGKPIELSGE 1139
Db 1239 EKLNRNEVTVLGE 1251
```

RESULT 8

```

MYS2_SCHPO
ID MYS2_SCHPO STANDARD; PRT: 1526 AA.
AC Q9US16; P78969;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE MYOSIN TYPE II HEAVY CHAIN 1.
GN MYO2 OR SPC645.05C.
OC Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A., AND FUNCTION.
RC STRAIN=972;
RX MEDLINE=98075862; PubMed=9415380;
RA May K.M., Watts F.Z., Jones N., Hyams J.S.;
RT "Type II myosin involved in cytokinesis in the fission yeast,
Schizosaccharomyces pombe.";
RL Cell Motil. Cytoskeleton 38:385-396(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Wood V., Rajandream M.A., Barrell B.G., Rieger M.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -! FUNCTION: REQUIRED FOR CELL DIVISION. IT IS A COMPONENT OF THE
CDC12 'SPOT', A STRUCTURE THOUGHT TO MARK THE SITE OF SEPTATION.
CC MAY WORK IN CONJUNCTION WITH MYO3.
CC -! SUBUNIT: BINDS TO CDC4 AND RLC1.
CC -! SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -! SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U75357; AAC49908.1;
CC HSP; AL049496; CAB39901.1;
CC HSP; P08799; LMND.
CC InterPro: IPR000048; IQ.
CC InterPro: IPR001609; myosin_head.
CC Pfam: PF00612; IQ.
CC Pfam: PF00063; myosin_head; 1.
CC PRINTS: PR00193; MYOSINHEAVY.
CC ProDom: PD000355; myosin_head; 1.
CC SMART; SM00015; IQ.
CC SMART; SM00242; MYSC; 1.
CC PROSITE; PS50096; IQ.
KW Myosin; Actin-binding; ATP-binding; Coiled coil; Calmodulin-binding;
KW Alkylation.
FT DOMAIN 1 757 MYOSIN HEAD-LIKE.
FT DOMAIN 758 787 IQ.
FT DOMAIN 787 1244 COILED COIL (POTENTIAL).
FT NP_BIND 170 177 ATP (POTENTIAL).
FT DOMAIN 634 656 ACTIN-BINDING (BY SIMILARITY).
FT DOMAIN 734 748 ACTIN-BINDING (BY SIMILARITY).
FT MOD_RES 674 674 ALKYLATION (BY SIMILARITY).
FT CONFLICT 1337 1337 S -> R (IN REF. 1).
SQ SEQUENCE 1526 AA; 176430 MW; D71D51D6578192BA CRC64;

Query Match 1.9%; Score 175.5; DB 1; Length 1526;
Best Local Similarity 19.9%; Pred. No. 0.0085;
Matches 230; Conservative 147; Mismatches 425; Indels 351; Gaps 54;

QY 39 FLHHLAQLVPEIYFAEMDPLEKQESVQMSIFTPLWYLFGE--PDI-CLEKLKHSKA 95
Db 473 FNHMFVLEQEYV-----KEIV-----WDFIDFGHDLQPTIDLEKANPIGI 516
```

Qy 96 FOLGRVFKSGTYYSCRD--C-----AIDPTCVLCMDFQDSVHKHNRHYKMHSTGGF- 148
Db 517 L-----SCLDECVMPKATDAFTSKLDALW--NKSLKYKPKFADQGEI 560
Qy 149 -----CDCG-DTEAWKTPFCVNHPEGRAGTIKNSRCPLNEEVIVQARKIPPSVIKYV 202
Db 561 LTHYAADVPYSTEGW-----LEKNTDPLNENV---AKLLAQSTNKHVA 600
Qy 203 EM-----TIWEEKELPPEL--OIREKNERY--CVLENDEH-- 235
Db 601 TLFSDYQETETKTVRGRTKGLFTVQARHKEQLNQLMNQNSTOPHRCIVNPEEKKM 660
Qy 236 HSYDHVYLSQALDCELAELAAQLHTTADKEGRRAVAGAYAAQCAEKEDIKSHSENVSO 295
Db 661 HTFNR-----PLVLGQLRCNGV-LEGIRITRAGF-----PNR 691
Qy 296 HPLHVEVLHSEIMAHQFALRLGSMNMKIM-----SYSDFRIFQOACLRPEPD 345
Db 692 LPFNDFRVRVEIMAHLPFTGYVESRRASVMIELBKIDEASVYRGVSKIFFKAGVLAEL 751
Qy 346 SENCLISRLMLDAKLYKGARKTLHELIFSSFFMEME-----YKLFAMEFVK 394
Db 752 ERRVATLQRLM---TMLQTRIRGELQRIKFIQKRLKDTQAIKLLQANLQVYNFEFTFPWAK 808
Qy 395 YKQLOXEYISDDHDSRISIT-----ALSVMFTVPTLARHLIEONVISVITETL 445
Db 809 LFFNLRLPILLSQNDKQKLRDABIEILKYLKQKQNSKSEVEDLVTNLSLAV-ENL 867
Qy 446 L-----EVLPEYLDL---NNKFNPGYSQDKLGRVYAVICDLKYLISKPTIW 490
Db 868 LTTERRAIALDKEELTRTOERLANIEDSFSTKQONENLQRESA---SLKQI----- 916
Qy 491 TERLMOFLEGFRFLKILTCMQMWEIRROVGQH-----IEVDPDWEAAIAIQMLKNIL 546
Db 917 NNELESELLE---KTSKVETLLSQNELKELSLDEKDLDTKGELESRLNENATVLSK 973
Qy 547 LMFOEWACADELLVAVYKCHKAVMRCSTSFSSSTVVGSCGSHSLTKSVRYVSEDLVS 606
Db 974 AEFNEQCKSQIETIVTKDAELDKL-----TKYISDYKTEIO-----EMLTKQKNEKSIQ 1024
Qy 607 IHLPLSRTLAGHLVRLSLGAV-SRLHEFVSFDFQVEVLVEYPLRLCIVLVAQVVAEMWR 665
Db 1025 QEGSLSESLEK---RVKRLERENSTLSDVSILKQKEEL-----SVLKGQVQELTI 1071
Qy 666 RNLGLISLQVFFYQ-DVYC-----REEMVDKDIIMLOIGASLMDPNKFLLLVLQRYEL 717
Db 1072 NN---LEEKVNYLEADYKQLPKLKELESNDKD-----OLYQL 1107
Qy 718 AEFNKTIKTDOLIKOYNTLIBEMLOVLIYIVGERVYVPGVGNVTKKEVTMRETIHLLC 777
Db 1108 QATNKELEAKVKCLNNIKSLTKELN-----KEKQCNLSASLKYIELQEI----- 1156
Qy 778 IEPHPSAIAKNLPENNETGLENVINKVATFKKPGVGHGVYELKDESLKDFNMVYFH 837
Db 1157 -----H-----ENLLKVSLENKKYKYGSL-QLDLEGLKQVDTNQE 1193
Qy 838 YSKTQ-----HSAKAEHMOK-KRRQENKD-----EALPPPPPEFCAPFSKVNL 881
Db 1194 LSKKHRLDTFNHESLLRQSASVYKELSLASSENKDLNKSLSLKQVNLSPKASKVPKL 1253
Qy 882 LNCDIMVILRTVPEAIDTDSNLWTEGMLQMAFHILALGLLEEKQLOKAPKEEVTFDF 941
Db 1254 -----ERKI-----TNLMHYSQUGKTFDEKRRAL 1279
Qy 942 YHKASRLGSSAMNTQMLEKLGIPQLEGQKDMITWILQMPDITVKRLREKSKCLIVATTS 1001
Db 1280 I--ASRNEELRSLLKSELESKRKL-EVEYQK-----VLEEVKTTLSRLSEVTLRLNKNVAD 1331
Qy 1002 SESIKNDIEIHDKKAERKKRKAARLHROKIMAQMSALQKNFTIETHKLMDYNTSEMPGK 1061
Db 1332 HESIRS-KLSEVENKLVDTRK-----ELNSALDSCKKREAEIHLRKEHRPS---GK 1378

Qy 1062 EDSIMEESTPAV 1074
Db 1379 ENNIPAVKTTTPV 1391
RESULT 9
ID CENE_HUMAN STANDARD; PRT; 2663 AA.
AC O02224;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CENTROMERIC PROTEIN E (CENP-E PROTEIN).
GN CENPE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93024922; PubMed=1406971;
RA Yen T.J., Li G., Schaar B.T., Sziliak L., Cleveland D.W.;
RT "CENP-E is a putative kinetochore motor that accumulates just before
RT mitosis.";
RL Nature 359:536-539(1992).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=95196755; PubMed=7889940;
RA Thrower D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;
RT "Mitotic HeLa cells contain a CENP-E-associated minus end-directed
RT microtubule motor.";
RL EMBO J. 14:918-926(1995).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=98437347; PubMed=9763420;
RA Chan G.K.T., Schaar B.T., Yen T.J.;
RT "Characterization of the kinetochore binding domain of CENP-E reveals
RT interactions with the kinetochore proteins CENP-F and hBUBR1.";
RL J. Cell Biol. 143:49-63(1998).
CC -!- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE
CC KINETOCORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE
CC OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT
CC AND/OR SPINDLE ELONGATION.
CC -!- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.
CC -!- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCORES DURING
CC CONGRESSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS
CC QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.
CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z15005; CAA78727.1; -
DR PIR; S28261; S28261.
DR HSSP; P03069; 1Z11.
DR MIN; 117143; -
DR InterPro; IPR001752; kinesin.
DR Pfam; PF00225; kinesin; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR PROSITE; PS00067; KINESIN_MOTOR_DOMAIN2; 1.
KW Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;
KW Cell cycle; Centromere.
FT DOMAIN 1 335
FT COILED COIL (POTENTIAL).
FT DOMAIN 336 2471
FT GLOBULAR (POTENTIAL).
FT NP_BIND 2472 2663
FT 86 93
FT ATP (BY SIMILARITY).

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE A KINASE ANCHOR PROTEIN 9 (PROTEIN KINASE A ANCHORING PROTEIN 9)

DE (PKRA9) (A-KINASE ANCHOR PROTEIN 450 KDA) (AKAP 450) (A-KINASE ANCHOR

DE PROTEIN 350 KDA) (AKAP 350) (HGAKAP 350) (AKAP 120 LIKE PROTEIN)

DE (HYPERION PROTEIN) (YOTIAO PROTEIN) (CENTROSOME- AND GOLGI-LOCALIZED

DE PKN-ASSOCIATED PROTEIN) (CG-NAP).

GN AKAP9 OR AKAP450 OR AKAP350 OR KIAA0803.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A. (ISOFORM 4).

RC TISSUE=Brain;

RX MEDLINE=98151389; PubMed=9482789;

RA Lin J.W., Wyszynski M., Madhavan R., Sealock R., Kim J.U., Sheng M.;

RT 'Yotiao, a novel protein of neuromuscular junction and brain that

RT interacts with specific splice variants of NMDA receptor subunit

RT NR1.1";

RL J. Neurosci. 18:2017-2027(1998).

RN [2]

RP SEQUENCE FROM N.A. (ISOFORM 2), AND VARIANT GLN-1347 INS.

RX MEDLINE=99219864; PubMed=10202149;

RA Witzak O., Skalihegg B.S., Kerker G., Bornens M., Tasken K.;

RA Jansen T., Oerstavik S.;

RT "Cloning and characterization of a cDNA encoding an A-kinase anchoring

RT protein located in the centrosome, AKAP450.1";

RL EMBO J. 18:1858-1868(1999).

RN [3]

RP SEQUENCE FROM N.A. (ISOFORM 3).

RC TISSUE=Brain;

RX MEDLINE=99287934; PubMed=10358086;

RA Takahashi M., Shibata H., Shimakawa M., Miyamoto M., Mukai H., Ono Y.;

RT "Characterization of a novel giant scaffolding protein, CG-NAP, that

RT anchors multiple signaling enzymes to centrosome and the golgi

RT apparatus";

RL J. Biol. Chem. 274:17267-17274(1999).

RN [4]

RP SEQUENCE FROM N.A. (ISOFORM 1).

RA Kemmer W.A., Deiss S., Schwarz U.;

RT "Cloning of Hyperion.1";

RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.

RN [5]

RP SEQUENCE OF 323-3911 FROM N.A. (ISOFORM 2).

RX MEDLINE=99115654; PubMed=9915845;

RA Schmidt P.H., Dransfield D.T., Claudio J.O., Hawley R.G.;

RA Trotter K.W., Milgram S.L., Goldenring J.R.;

RT "AKAP350, a multiply spliced protein kinase A-anchoring protein

RT associated with centrosomes.1";

RL J. Biol. Chem. 274:3055-3066(1999).

RN [6]

RP SEQUENCE OF 1802-3876 FROM N.A. (ISOFORM 5).

RC TISSUE=Lymphoblast;

RA Hinds K., Sutterer C., Becker M., Hawkins M.;

RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.

RN [7]

RP SEQUENCE OF 2157-3911 FROM N.A. (ISOFORM 6).

RC TISSUE=Lung;

RA Milgram S.L., Goldenring J.R., Schmidt P.H.;

RT "AKAP350: A multiply spliced family of proteins with centrosomal

RT association.1";

RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.

RN [8]

RP SEQUENCE OF 2212-3911 FROM N.A. (ISOFORM 2/3).

RC TISSUE=Brain;

RX MEDLINE=99087487; PubMed=9872452;

RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.;

RA Tanaka A., Kotani H., Nomura N., Ohara O.;

RT "Prediction of the coding sequences of unidentified human genes. XI.

RT The complete sequences of 100 new cDNA clones from brain which code

RT for large proteins in vitro.1";

RL DNA Res. 5:277-286(1998).

RN [9]

RP SEQUENCE OF 17-1800 FROM N.A.

RA Wu X., Graves T., Bradshaw H.;

RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: BINDS TO TYPE II REGULATORY SUBUNITS OF PROTEIN KINASE

CC A. SCAFFOLDING PROTEIN THAT ASSEMBLES SEVERAL PROTEIN KINASES AND

CC PHOSPHATASES ON CENTROSOME AND GOLGI APPARATUS WHERE PHYSIOLOGICAL

CC EVENTS CAN BE REGULATED BY PHOSPHORYLATION STATE OF PROTEIN

CC SUBSTRATES. ISOFORM 4/YOTIAO IS ASSOCIATED WITH THE N-METHYL-D-

CC ASPARTATE RECEPTOR AND IS SPECIFICALLY FOUND IN THE NEUROMUSCULAR

CC JUNCTION (NMJ) AS WELL AS IN NEURONAL SYNAPSES EXPLAINING THAT ITS

CC ROLE MAY BE TO ORGANIZE POSTSYNAPTIC SPECIALIZATIONS.

CC -!- SUBUNIT: INTERACTS WITH THE REGULATORY REGION OF PROTEIN KINASE N

CC (PKN), PROTEIN PHOSPHATASE 2A (PP2A), PROTEIN PHOSPHATASE 1 (PP1)

CC AND THE IMMATURE NON-PHOSPHORYLATED FORM OF PKC EPSILON.

CC -!- SUBCELLULAR LOCATION: CENTROSOMAL IN MANY CELL TYPES AND

CC CYTOPLASMIC IN PARIETAL CELLS.

CC -!- ALTERNATIVE PRODUCTS: 6 ISOFORMS; 1 (SHOWN HERE), 2, 3/CG-

CC NAP, 4/YOTIAO, 5 AND 6/AKAP350; ARE PRODUCED BY ALTERNATIVE

CC SPLICING.

CC -!- TISSUE SPECIFICITY: WIDELY EXPRESSED. ISOFORM 4/YOTIAO IS HIGHLY

CC EXPRESSED IN SKELETAL MUSCLE AND IN PANCREAS.

CC -!- DOMAIN: RII BINDING SITE, PREDICTED TO FORM AN AMPHIPATHIC HELIX,

CC COULD PARTICIPATE IN PROTEIN-PROTEIN INTERACTIONS WITH A

CC COMPLEMENTARY SURFACE ON THE R-SUBUNIT DIMER.

CC -!- CAUTION: REF.6 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO TWO

CC FRAMESHIFTS IN POSITIONS 3782 AND 3811.

CC -!- CAUTION: REF.9 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO FOUR

CC FRAMESHIFTS IN POSITIONS 29, 1653, 1699 AND 1735.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; AJ131693; CAB40713.1; -

CC EMBL; AB019691; BAA78718.1; -

CC EMBL; AJ010770; CAA09361.1; -

CC EMBL; AF026245; AAB86384.1; -

CC EMBL; AF083037; AAD22767.1; -

CC EMBL; AC004013; AAB96867.1; ALT_FRAME.

CC EMBL; AF091711; AAD39719.1; -

CC EMBL; AB018346; BAA34523.1; -

CC EMBL; AC000066; AAC60380.1; ALT_FRAME.

CC MIM; 604001; -

CC Coiled coil; Alternative splicing; Polymorphism.

KW DOMAIN 2554 2567 PKA-RII SUBUNIT BINDING DOMAIN.

FT DOMAIN 164 914 COILED COIL (POTENTIAL).

FT DOMAIN 944 1022 COILED COIL (POTENTIAL).

FT DOMAIN 1100 1185 COILED COIL (POTENTIAL).

FT DOMAIN 1253 1280 COILED COIL (POTENTIAL).

FT DOMAIN 1336 1392 COILED COIL (POTENTIAL).

FT DOMAIN 1434 1459 COILED COIL (POTENTIAL).

FT DOMAIN 1585 1659 COILED COIL (POTENTIAL).

FT DOMAIN 1857 2455 COILED COIL (POTENTIAL).

FT DOMAIN 2544 2561 COILED COIL (POTENTIAL).

FT DOMAIN 2603 2776 COILED COIL (POTENTIAL).

FT DOMAIN 3065 3092 COILED COIL (POTENTIAL).

FT DOMAIN 3124 3470 COILED COIL (POTENTIAL).

FT DOMAIN 3587 3689 COILED COIL (POTENTIAL).

FT DOMAIN 3726 3730 POLY-LEU.

FT DOMAIN 203 292 GLN-RICH.

FT DOMAIN 321 1010 GLU-RICH.

FT DOMAIN 1846 2772 GLU-RICH.

FT VARSPPLIC 17 28 MISSING (IN ISOFORM 2 AND ISOFORM 3).

FT VARSPPLIC 1637 1642 OLOEEI -> LATERR (IN ISOFORM 4).

FT VARSPPLIC 1643 3911 MISSING (IN ISOFORM 4).

FT VARSPPLIC 2175 2182 MISSING (IN ISOFORM 3).

FT VARSPPLIC 2175 2183 SADTFQKVE -> Q (IN ISOFORM 6).

FT VARSPPLIC 2895 2907 VGFYNNCFSTLC -> GSSIPELAHSDAYQTREICSS


```

Db      1359  -----PKEHLWQFL-----CIIFESQVLEQVPQKYSTDK-----SRLEDA 1394
Qy      1548  TNLFLEQYWDYTRPLLRWCADP--ALLNCLKOKNTVVRYPRKRNLSL 1594
Db      1395  RELTLMFEE-----DPTVAIQTCIRLLDYLAKLPAKSSSL 1428

RESULT 13
MY5A_MOUSE
ID AC MY5A_MOUSE STANDARD; PRT; 1853 AA.
AC Q99104:
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MYOSIN VA (MYOSIN 5A) (DILUTE MYOSIN HEAVY CHAIN, NON-MUSCLE).
GN MYOSA OR DILUPE.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
ON NCBI_TaxID=10090;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RC MEDLINE=91141583; PubMed=1996138;
RA Mercer J.A., Seperack P.K., Strobel M.C., Copeland N.G.,
RA Jenkins N.A.;
RA "Novel myosin heavy chain encoded by murine dilute coat colour
RT locus".
RL Nature 349:709-712(1991).
RN [2]
RP REVISIONS.
RA Mercer J.A., Seperack P.K., Strobel M.C., Copeland N.G.,
RA Jenkins N.A.;
RA Nature 352:547-547(1991).
CC -!- FUNCTION: MAY BE INVOLVED IN MELANOSOME TRANSPORT, OR
CC INVOLVED IN DENDRITE FORMATION.
CC -!- SUBUNIT: MAY BE A HOMODIMER, WHICH ASSOCIATES WITH MULTIPLE
CC CALMODULIN OR MYOSIN LIGHT CHAINS.
CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 DILUTE DOMAIN.
CC -!- SIMILARITY: CONTAINS 6 IQ DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no
CC modified and this statement is not removed. Usage by and for commer-
CC ential entities requires a license agreement.(See http://www.isb-sib.ch/announc-
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X57377; CAA040651.1; -.
CC PIR; A46761; A46761.
CC DR HSP; P08799; 1MND.
CC DR MGD; MG1:105976; Myo5A.
CC DR InterPro; IPR002710; DfL.
CC DR InterPro; IPR000048; IQ.
CC DR InterPro; IPR001609; myosin_head.
CC DR Pfam; PF01843; DfL; 1.
CC DR Pfam; PF00612; IQ; 6.
CC DR Pfam; PF00063; myosin_head; 1.
CC DR PRINTS; PR00193; MYOSINHEAVY.
CC DR ProDom; PD0000355; myosin_head; 1.
CC DR ProDom; PD003376; DfL; 1.
CC DR SMART; SM00015; IQ; 6.
CC DR SMART; SM00242; MYSC; 1.
CC DR PROSITE; PS50096; IQ; 6.
CC DR
CC KW Myosin; Repeat; ATP-binding; Calmodulin-binding; Actin-binding;
CC Coiled coil; Phosphorylation.
CC FT DOMAIN 1 765 MYOSIN HEAD-LIKE.
CC FT DOMAIN 766 788 IQ 1.
CC FT DOMAIN 789 813 IQ 2.
CC FT

```

```
QY 970 GOKDMITWLOFDTVKRLREKSLIVATTSGSESIKNDKIDTHDKKAEKRKAARLUH 1029
Db 1167 QEKQL-----MODELDR-----KEQVFRSKAKKEERQPIRGAEL- 1201
QY 1030 RQKIMAOMASALQKNFIE-HKLMYDNTSEMPGKEDSIMEESTPAVSDYSRIALGPKRGP 1088
Db 1202 -----EYESLKQLESENKKLNKLNEL-----RKALSEKSAP 1235
QY 1089 SVTE-----KEVLRCILCOEPOEYKIENNAMVLSACVOKSTALTQHRGKPIELSGE 1139
Db 1236 EVTAPGAPAYRVLMEQLTSV--SEELDVKEEVLTLRSOLVSQKEAI-----QPKDDKNT 1288
QY 1140 ALDPLFMDPDLAYGTYTSCGHVHACWQYFEAVQLSSQORIHVDLFDLESGEYLCPL 1199
Db 1289 MTDSTILLEDQVKMKDKGEIAQAYTGLKTRNLLESQLSQKSRSHNEAEALRGE----- 1343
QY 1200 CKSLCNTVIPIPLQPOKINSADALALLTLARWIOTVLARISG---YNIRHAKGENP 1256
Db 1344 IQSL-----KEENRQQLLAQNQL-----PPEARIEASLOHEITRLTNEN- 1385
QY 1257 IPIFFNOGMGSTLEFHSI-----LSFGVESSIKYSINSIKEMVILFATTI--YRI 1304
Db 1386 --LYFEELYADPKKYQYRISLYKRMIDLMQLEKQDKTVRKLLKQKLVFAKKIGELEV 1443
QY 1305 G-----LKVPDPDERD----- 1314
Db 1444 GOMENISPGQITDEPIRPVNIPIRKEKDFQGMLEYKREDEQKLVKNLILELKPGRGVAVNLI 1503
QY 1315 PRVPLMTWSTC-----AFTIOATENLL--GDGKPLFGALQN----- 1349
Db 1504 PGLPAYILFMCVRHADYLNDDQKVRSLTSTINSIKVLKRGDDFETVSWLSTWCREL 1563
QY 1350 ---RQHGLKALMOPAVQAQRITPCOVLIQKHLVRLLSVVLNPKSEDTPCLLSIDLF--- 1403
Db 1564 HCLKQYSGEEGFM-----KH-----NTSQNEHCLTNFDLAEVR 1597
QY 1404 HVLVG-AVLAPFSLWDDPVDLPQSSVSSYNHLYFLHITMAHMLQILLTV-DTGLPLA 1461
Db 1598 QVLSDLAIQIQQLVRVLENILQPMIVSGMLEH-----ETIQGVSGVKPTGL--- 1644
QY 1462 QVQEDSEEAHSASSFAETISQVTSIGCDIPGWVLYWSLKNITPYL-RCAALFFHYLL 1520
Db 1645 -----RRTSIADEGTYTLDLSILQLNSFHS-VMOHGMPDELKQVVKQMFYIV 1694
QY 1521 GVTTPPEELHTNSAEYSA LCSY--LSIPTNLFLLFQEWYDVTVPRLQWCADPALLNC 1577
Db 1695 GAITLNNLLR-----KDMCSWSKGMQIRNV-----SOLEEWLRDKNLMNS 1736
QY 1578 LKQKNTVYRPRKNSLIELPDYSCLLNQASHFRCPRSADDERKHPVLCFLFCGAILCSQ 1637
Db 1737 -GAKETL-----EPLIQAAQLLQVKKTKTDDAE--ATCSMCNALTTAQ 1776

RESULT 14
MY5A_HUMAN STANDARD; PRT; 1855 AA.
AC Q9Y4I1; OSUE30; Q9UE31; Q07902; Q16249; O60653;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE MYOSIN VA (MYOSIN 5A) (DILUTE MYOSIN HEAVY CHAIN, NON-MUSCLE)
DE MYOSIN HEAVY CHAIN 12) (MYOXIN).
GN MYO5A OR MYH12.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=Skin;
RA Meurers B.H., Zimmermann R., Vosberg H.P.;
RT "The complete cDNA for human myosin heavy chain 12, a class V
myosin.";
```

```
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1), AND VARIANT CVS-1246.
RX MEDLINE=97351514; PubMed=9207796;
RA Pastural E., Barrat F.J., Dufourcq-Lagelouse R., Certain S., Sanal O.,
RA Jabado N., Seger R., Griscelli C., Fischer A., de Saint Basile G.;
RT "Griscelli disease maps to chromosome 15q21 and is associated with
mutations in the myosin-Va gene.";
RL Nat. Genet. 16:289-292(1997).
RN [3]
RP ERRATUM.
RA Pastural E., Barrat F.J., Dufourcq-Lagelouse R., Certain S., Sanal O.,
RA Jabado N., Seger R., Griscelli C., Fischer A., de Saint Basile G.;
RL Nat. Genet. 23:373-373(1999).
RN [4]
RP SEQUENCE OF 638-1477 FROM N.A. (ISOFORM 2).
RC TISSUE=Fetal brain;
RX MEDLINE=94245227; PubMed=8188282;
RA Engle L.J., Kennett R.H.;
RT "Cloning, analysis, and chromosomal localization of myoxin (MYH12),
the human homologue to the mouse dilute gene.";
RL Genomics 19:407-416(1994).
RN [5]
RP SEQUENCE OF 1061-1498 FROM N.A. (ISOFORM 2).
RC TISSUE=Brain;
RX MEDLINE=95136715; PubMed=7835087;
RA Moore K.J., Testa J.R., Francke U., Milatovich A., Copeland N.G.,
RA Jenkins N.A.;
RT "Cloning and regional assignment of the human myosin heavy chain 12
(MYH12) gene to chromosome band 15q21.";
RL Cytogenet. Cell Genet. 69:53-58(1995).
RN [6]
RP SEQUENCE OF 1358-1460 FROM N.A. (ISOFORM 3).
RA Edgar A.J., Bennett J.P.;
RT "Inhibition of dendrite formation in melanocytes transiently
transfected with antisense DNA to myosin V.";
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [7]
RP FUNCTION.
RX MEDLINE=99376094; PubMed=10448864;
RA Mehta A.D., Rock R.S., Rief M., Spudich J.A., Mooseker M.S.,
RA Cheney R.E.;
RT "Myosin-V is a processive actin-based motor.";
RL Nature 400:590-593(1999).
CC -!- FUNCTION: PROGRESSIVE ACTIN-BASED MOTOR THAT CAN MOVE IN LARGE
STEPS APPROXIMATING THE 36-NM PSEUDO-REPEAT OF THE ACTIN FILAMENT.
MAY BE INVOLVED IN MELANOSOME TRANSPORT, OR ALTERNATIVELY, IT MAY
BE REQUIRED FOR SOME POLARIZATION PROCESS INVOLVED IN DENDRITE
FORMATION.
CC -!- SUBUNIT: MAY BE A HOMODIMER, WHICH ASSOCIATES WITH MULTIPLE
CALMODULIN OR MYOSIN LIGHT CHAINS.
CC -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS: 1 (SHOWN HERE), 2 AND 3; ARE
PRODUCED BY ALTERNATIVE SPLICING.
CC -!- DISEASE: DEFECTS IN MYO5A ARE A CAUSE OF GRISCELLI SYNDROME (GS).
GS IS A RARE AUTOSOMAL RECESSIVE DISORDER THAT RESULTS IN
PIGMENTARY DILUTION OF THE SKIN AND HAIR, THE PRESENCE OF LARGE
CLUMPS OF PIGMENT IN HAIR SHAFTS, AND AN ACCUMULATION OF
MELANOSOMES IN MELANOCYTES. MOST PATIENTS ALSO DEVELOP AN
UNCONTROLLED T LYMPHOCYTE AND MACROPHAGE ACTIVATION SYNDROME,
KNOWN AS HEMOPHAGOCYTIC SYNDROME, LEADING TO DEATH IN THE ABSENCE
OF BONE MARROW TRANSPLANTATION.
CC -!- SIMILARITY: CONTAINS 6 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 DILUTE DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
```


QY	1221	ENADALAQI-----LTLRWIQTVLARI	SGYNIRHAKGNPIPIFFNOCMGDSTL---	1270
			: : : : : :	
Db	1411	ENLDLMEQLEKQDKTVRKLLKQKLVFAKKI	GEVYGOENISPGQIIT-----DEPIRPV	1464
QY	1271	-----EPHSITLSFGVSSSTK-----	-----	1288
Db	1465	NTPREKDFQGMLEYKKEDEQKLVKNLLEL	KPRGVAVNLPGLPAVILPWCVRHADYLN	1524
QY	1289	SIKEMVILFATTIYRIGLKVPPDERDPRV	MLTW--STCAFTIOAIENLLGDGKPLFGA	1346
Db	1525	DOQVRSLLTSPINSI-KKVLKRGKDDFET	YSEWLSNTRCF-LHCLKQYSGEEGFMKHN	1582
QY	1347	LQNRQHNGLKALMQFAVA---QRITCP	OVILQKHLVRLLSVVLNPIKSEDTPCLL	1403
Db	1583	SRQNEH---CLTNFDLAEYRQVLSDAIQ	YVQQLVR-----VLENI-----	1620
QY	1404	HVLGVNAFLAPSLYWDPDVLPDPSSV	SSYNHLLYFLHLITMAHMLQILLTV-D	1462
Db	1621	-----	LPMTIVSGMLH-----ETIQGVSGVK	1646
QY	1463	VQEDSEAHSSAFEAISQVTSGGSGCDI	PCWYLWVSLKNGIIPYL-RCAALFFHYLL	1521
Db	1647	-----AKRTSSIADEGTIVLDSIL	RQLNSFHS-VMCQHGMDPELIKQVVKOM	1697
QY	1522	VTPPEELHTNSAEGEYSALCSY---LS	LPTNLFLFQEVYDVRPLLRMCADPALL	1578
Db	1698	AITLNNLLR-----KDMCSWSKMGQI	RYNV-----SQLEEWLRDKNLMS-	1738
QY	1579	KOKNTVYVYPRKRNLSIELPDDYSC	LLNQASHFRCPSPRADERKHPVLCLF	1637
Db	1739	GAKETL-----EPLIOAAQLQVKK	KTDDAE--AICSMCNALTTAQ	1778
RESULT 15				
EVPL_HUMAN				
ID	EVPL_HUMAN	STANDARD;	PRT; 2033 AA.	
AC	Q92817;			
DT	20-AUG-2001	(Rel. 40, Created)		
DT	20-AUG-2001	(Rel. 40, Last sequence update)		
DT	20-AUG-2001	(Rel. 40, Last annotation update)		
DE	ENVOPLAKIN (210 KDA PARANEOPLASTIC	PEMPHIGUS ANTIGEN) (P210) (210 KDA		
DE	CORNIIFIED ENVELOPE PRECURSOR).			
GN	EVPL.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata;	Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini;	Hominidae; Homo.		
OX	NCBI_TaxID=9606;			
ON	[1]			
RP	SEQUENCE FROM N.A.			
RP	TISSUE=Keratinocytes;			
RX	MEDLINE=96326676; PubMed=8707850;			
RA	Ruhrberg C., Hajibagheri M.A.N., Simon M.,	Dooley T.P., Watt F.M.;		
RA	Evans K.E., Ellis A., Watt F.M., Bishop D.T.,	Spurr N.K.,		
RA	Stevens H.P., Leigh I.M., Reis A., Kelsell D.P.,	Field J.K.;		
RT	"Envoplakin, a possible candidate gene for	focal NEPK/esophageal		
RT	cancer (TOC): the integration of genetic and	physical maps of the TOC		
RT	region on 17q25.,"			
RL	Genomics 59:234-242(1999).			
CC	!- FUNCTION: COMPONENT OF THE CORNIIFIED	ENVELOPE OF KERATINOCYTES.		
CC	MAY LINK THE CORNIIFIED ENVELOPE TO	DESMOSOMES AND INTERMEDIATE		
CC	FILAMENTS.			
CC	!- SUBUNIT: MAY FORM A HOMODIMER OR A	HETERODIMER WITH PPL.		
CC	!- SUBCELLULAR LOCATION: COLOCALIZED	WITH DSP AT DESMOSOMES AND ALONG		
CC	INTERMEDIATE FILAMENTS.			
CC	!- TISSUE SPECIFICITY: EXCLUSIVELY	EXPRESSED IN STRATIFIED SQUAMOUS		

```

CC      EPIHELIX.
CC      -1- INDUCTION: DURING DIFFERENTIATION OF EPIDERMAL KERATINOCYTES.
CC      -1- SIMILARITY: CONTAINS 7 PLECTIN REPEATS.
CC      -1- SIMILARITY: CONTAINS 1 SPECTRIN REPEAT.
CC      -1- SIMILARITY: BELONGS TO THE PLAKIN OR CYTOLINKER FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      ENBL; U53786; AAC64662.1; -.
CC      EMBL; U72849; AAD00186.1; -.
CC      EMBL; U72843; AAD00186.1; JOINED.
CC      DR EMBL; U72845; AAD00186.1; JOINED.
CC      DR EMBL; U72846; AAD00186.1; JOINED.
CC      DR EMBL; U72847; AAD00186.1; JOINED.
CC      DR EMBL; U72848; AAD00186.1; JOINED.
CC      DR EMBL; U72848; AAD00186.1; JOINED.
CC      MIM; 601590; -.
CC      InterPro: IPR001101; plectin_repeat.
CC      InterPro: IPR002017; Spectrin.
CC      DR Pfam; PF00681; Plectin_repeat; 3.
CC      DR SMART; SM00250; PLEC; 7.
CC      DR SMART; SM00150; SPEC; 1.
CC      KW Repeat; Coiled coil; Cytoskeleton; Structural protein.
CC      FT DOMAIN 1 841 GLOBULAR 1.
CC      FT DOMAIN 842 1673 CENTRAL FIBROUS ROD DOMAIN.
CC      FT DOMAIN 1674 2033 GLOBULAR 2.
CC      FT DOMAIN 12 28 4 X 4 AA TANDEM REPEATS OF K-G-S-P.
CC      FT REPEAT 229 330 SPECTRIN.
CC      FT DOMAIN 845 1135 COILED COIL (POTENTIAL).
CC      FT REPEAT 1185 1226 PLECTIN 1.
CC      FT REPEAT 1678 1713 PLECTIN 2.
CC      FT REPEAT 1818 1855 PLECTIN 3.
CC      FT REPEAT 1856 1893 PLECTIN 4.
CC      FT REPEAT 1894 1931 PLECTIN 5.
CC      FT REPEAT 1932 1969 PLECTIN 6.
CC      FT REPEAT 1970 2007 PLECTIN 7.
CC      SQ SEQUENCE 2033 AA; 231616 MW; B8DC6E2B52221938 CRC64;
Query Match 1.7%; Score 158; DB 1; Length 2033;
Best Local Similarity 19.3%; Pred. No. 0.16;
Matches 273; Conservative 212; Mismatches 509; Indels 420; Gaps 63;

Qy 195 PSVIKYVYVMTWEERKELPELQIRKERNYYCVLFNDEHSHVDHYISLQRAL--DCEL 253
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 780 PSQIAYKL-----QAQRKLQEQISRR-----DRATASHLSQALQAALQDVEL 823

Qy 254 AEQLHTTADKERRAVKAGAAACQAEKEDIKSHSNYSQHPHVEVLHSEIMAHQKF 313
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 824 --QADTYRCSLPTLAVSAPKRPVRAPLOEQTAQEKNLAKAYTEVAAAQQQLLOQLF 880

Qy 314 ALRUGSNWKNIMSYSSDFRQIFQACLRPSDSENPCLISR-----LMLWDAKLYKGARKI 369
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 881 ARK----MLKKELSEDIRTH-----DAKQGESPAQAAGRESEALKAQLEERKRVARV 931

Qy 370 LHELIFSFTMEMEYKKLFAMEFYKKYKQLQKEYISDDHRSITSITALSVQMTFVPTLAR 429
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 932 QHEL-----EAQRSOLLQRTQRPRLERKEEV-----VEFYRDPQLEG 970

Qy 430 HL-----IEQNVISVITETLLEVLPEYLDNRNKNFQGYSDQKLGRRVAVICDLKYL 483
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 971 SLSRVKAQVEEGKRRAGLQADLEV-----AAQKVQLESKRKTMOPHL 1014

Qy 484 ISKPTIWT-----LRWF-----LEGFRSLKILTCMGMEIRR 520
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 1015 LTRKVTQVQDPLGDSQAQRLRQIQOLRGEDAVISARLEGLKKELLALBKREVDVVKEV 1074

Qy 521 QVGQHIEVDPDWE---AAIAIQMLKNILMLFQEWACDDEL--LLVAYKECHKAVMRC5 575

```

Search completed: February 28, 2002, 10:06:39
Job time: 275 sec

```

Db 1075 VKEVWVKNLEWVRAAQAQLRQMEDAARRKO---AEEAVAKQARIEDLERAISVE 1131
Qy 576 TS-FISSKTVQSCGHSLETKSYR-VSEDLVSIHLPLSRITLAGLHVRLS-----RLG 626
Db 1132 PKVIVKEVKVQEDPGLLQESSRLSLLEERTKNATLARELSDLHSKYSVVEKQRPVQ 1191
Qy 627 AVSRLHEFVSFEDFQVEVLVEYPLRCLVLVAQVVAEMWRNG-----LSLI 672
Db 1192 LQERVHEI-----FOVDPETEQEITRLKAKLOEMA--GKRSQVEKEVEKLLPDLEVLRAQ 1244
Qy 673 SOVFYQDV-----KCREEMVXDIIIMQIGASLMDPNKF--LLVLQ-----RYE 716
Db 1245 KPTVEYKEVQEVVRHERSPVLEIRDLRAQLNELVNSHGRSQOQLIRLQGERDEWRRE 1304
Qy 717 LAEAFNKTIS-----TKDQDLIKOYNTLIEMLQV-----LIY-----IVGERY 755
Db 1305 RAKVETKTVSKVVRHEKDPVLEKEAERLQEVREAAQKRAAEDAVYELQSKRLLLERR 1364
Qy 756 VPGVGNVTKEEVPMREIHLICIEPM---PHSAIAKNLPENENNENETGLENVINKVATFK 812
Db 1365 KP-----EKKVVVQEVV-VTKDPKLRREHSRLSGSLDEEVGRRRQLELEVQOL---R 1413
Qy 813 PGVSGHGVYELKDESKDFNMVYHYHSKTOHSAEHMOKKARKKOENKDEALPPPPPEFC 872
Db 1414 AGVE-----EOEGLLSF-----QEDRSKKLAVERELQTLRIQIELEKRP-- 1454
Qy 873 PAFSKVINLNCIMMIVILRTVFERAI-DPDSNLWTGEMLOMAFHILALGLLEKQOLQK 931
Db 1455 -----TVOEKIIMEEVVKLEKDPDLEKSTEALRWDLDOEQTKQVTE 1494
Qy 932 APEE-----EVTDFYHKASRLGSSAMNIQMLEKLGIPLEGOKDMITWILQMFDTVKR 987
Db 1495 LNRECKNLQVQIDVLOKAK---SQEXTIYKEVIRQVDRVLEDERARVWENLNRPTAQ 1551
Qy 988 LREKSC-----LIVATTSG-SESINDEITHDKKAER---KKAAEARLHROKIMQAM 1037
Db 1552 ARBEEARRLERIDRAETLGRWTSRESELOARQADQOCGRLOQELRALEROK---QQ 1608
Qy 1038 SALQKNFIETHKLMYDNTSEMPCKEDSIMEEESTPAVSDYSRIALGPKRGPSTEKEVLT 1097
Db 1609 QTLQLO--EESKLLSOKT-----ESEROKAA-----ORGOELSRLE--A 1643
Qy 1098 CILCOERQEVKIENNAMVACSQVKS--TALTQHRGKPIELSGEALDP---LPMDDPLAY 1152
Db 1644 AILREKQIYEKERTLRLHAKYSRELSQETQOTRETNLSTKISILEPETCKDMSPEAY 1703
Qy 1153 -----GTY-----TGSCGHVMHVCWQKYFEAVQLSSQORIHVDLFDL 1190
Db 1704 KRGIIDRGVYLOLQLECEWEEVTTSGPCE-----ESV-----LLDR 1741
Qy 1191 ESCEYLCPLCKSLCNTVPIPILOPQKINSENADALAQLLTARWQTVLARISGYNIRH 1250
Db 1742 KSGQYSIEAALRC-----RRISKEE-----YHL-Y 1766
Qy 1251 AKGENPIPIFNQMGD-----STLEFHSILSGFVSSIKYSNIKEMVILFATTIYRIGL 1306
Db 1767 KDGHLPISFALLVAGETKPFSSLSIGSIIS----- 1797
Qy 1307 KVPDDRPRVPMLTWSTCAFTIQAENLIGDECKPLFGALQNRHNGKALMQFAVAQR 1366
Db 1798 -----KSPLASAPQSTSFPS-PSFSLGLGDDSPFIAGIYDTTDTN--KCSIKTAVAKN 1848
Qy 1367 ITCP---QVLIQKH-----LVRLLSVVLNINKSEDPCLLSIDLHVLVGVAVLAFPSLYW 1418
Db 1849 MLDPIITGOKLLEAQAATGGIVDILLSRERYSVHKAMERGLIENTSTQRLNNAQKFTGI-- 1906
Qy 1419 DDPVDLQPSVSSVSSYN-----HLYLPHL 1441
Db 1907 EDPVTKRLSVGEAVOKGWMMPRESVLPHLQVQHL 1940

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 28, 2002, 09:58:03 ; Search time 28.88 Seconds
(without alignments)
4613.203 Million cell updates/sec

Title: US-09-724-126A-2

Perfect score: 9224

Sequence: 1 MADEEAGGTERMEISAEPLQ.....ETIARQETNQLFGFNWOLL 1749

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_68:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	8617	93.4	1757	T14318	ubiquitin-protein
2	1945.5	21.1	1927	T25604	hypothetical prote
3	1038	11.3	1958	T39808	hypothetical prote
4	914	9.9	2052	T37711	probable n-end-rec
5	689.5	7.5	1950	S12332	ubiquitin--protein
6	647.5	7.0	1941	T30554	ubiquitin-protein
7	606.5	6.6	1225	T48251	ubiquitin-protein
8	549.5	6.0	1872	S64851	probable membrane
9	343	3.7	795	T48252	ecceiferum3 (CER3)
10	218	2.4	271	T40238	hypothetical prote
11	205.5	2.2	3187	JC5637	364K Golgi complex
12	190.5	2.1	3259	A56539	giantin - human
13	186	2.0	1410	A57013	early endosome ant
14	182	2.0	2442	T08621	centrosome associa
15	180.5	2.0	1413	T26467	hypothetical prote
16	180	2.0	1830	S19188	myosin-V - chicken
17	179.5	1.9	2346	T13829	Tpr homolog - frui
18	176	1.9	3225	I52300	giantin - human
19	175.5	1.9	1427	S22695	restin - human
20	175.5	1.9	1326	T41522	myosin ii - fission
21	175.5	1.9	2663	S28261	centromere protein
22	174.5	1.9	1392	A43336	microtubule-vesicl
23	172.5	1.9	1780	T17272	hypothetical prote
24	168	1.8	1837	T41023	probable nuclear p
25	166	1.8	1676	E71410	probable centromer
26	164.5	1.8	1381	T01799	hypothetical prote
27	164	1.8	2954	T14156	kinesin-related pr
28	163.5	1.8	1339	H81307	restriction modifi
29	163	1.8	1853	A46761	myosin heavy chain

ALIGNMENTS

RESULT 1

T14318

ubiquitin-protein ligase E3-alpha - mouse

N:Alternate names: N-recognin E3-alpha

C:Species: Mus musculus (house mouse)

C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

C:Accession: T14318

R:Kwon, Y.T.; Reiss, Y.; Fried, V.A.; Hershko, A.; Yoon, J.K.; Gonda, D.K.; Sangani, P.

Proc. Natl. Acad. Sci. U.S.A. 95, 7898-7903, 1998

A:Title: The mouse and human genes encoding the recognition component of the N-end ru

A:Reference number: Z17977; MUID:98318583

A:Accession: T14318

A:Status: preliminary; translated from GB/EMBL/DBBJ

A:Molecule type: mRNA

A:Residues: 1-1757 <KWO>

A:Cross-references: EMBL:AF061555; NID:g3170886; PID:g3170887; PIDN:AAC40165.1

C:Genetics:

A:Gene: Ubql

A:Map position: 2

Query Match 93.4%; Score 8617; DB 2; Length 1757;
Best Local Similarity 92.1%; Pred. No. 0;
Matches 1619; Conservative 64; Mismatches 66; Indels 8; Gaps 3;

Qy 1 MADEEAGGTERMEISAEPLQTPORLASWDDQVDFYTAFLHLAQLVPEIYFAEMDPDLE 60

Db 1 MADEEMDGAERMDVSPPEPLAPQRPASWDDQVDFYTAFLHLAQLVPEIYFAEMDPDLE 60

Qy 61 KOEEVQMSIFTPLEWYLFEGDPDLCLEKLHSGAFQLCGRVFKSGETTYSCRDCAIDPT 120

Db 61 KOEEVQMSIFTPLEWYLFEGDPDLCLEKLHSGAFQLCGRVFKSGETTYSCRDCAIDPT 120

Qy 121 CVLCHDCFDQSDVHKHRYKMHHTSTGGFCDCGDTGTEAWKTGPFVCHNHEPGRAGTIKENSRC 180

Db 121 CVLCHDCFDQSDVHKHRYKMHHTSTGGFCDCGDTGTEAWKTGPFVCHNHEPGRAGTIKENSRC 180

Qy 181 PLNEEVIVQARKIFPSVIKYVEMTITWEEKELPPELQIREKNERYCYVLFNDEHHSYDH 240

Db 181 PLNEEVIVQARKIFPSVIKYVEMTITWEEKELPPELQIREKNERYCYVLFNDEHHSYDH 240

Qy 241 VIYSLQALDCLAEALHTTAIDKEGRVAKAGAAACQAEKEDIKSHSENVSQHPLHV 300

Db 241 VIYSLQALDCLAEALHTTAIDKEGRVAKAGAAACQAEKEDIKSHSENVSQHPLHV 300

Qy 301 EVLHSEVMAHQFALRLGSGWMNKIMSYSDSDFRQICQACLRERPDSENPCLSRLMLWDA 360

Db 301 EVLHSEVMAHQFALRLGSGWMNKIMSYSDSDFRQICQACLRERPDSENPCLSRLMLWDA 360

Qy 361 KLYGKARKILHELIFSSFFMEMEYKKLFAMEFVKYKQLOKEYISDDHRSISITALSQ 420

Db 361 KLYGKARKILHELIFSSFFMEMEYKKLFAMEFVKYKQLOKEYISDDHRSISITALSQ 420

bullous pemphigoid
probable myosin he
NMDA receptor-bind
hypothetical prote
transport protein
toxin A - Clostrid
reticulocyte-bind
1-phosphatidylinos
hypothetical prote
1-phosphatidylinos
myosin heavy chain
rhoetry protein -
dystrophin, muscle
MSP-300 protein -
hypothetical prote
probable DNA repai

QY 217 LQI--REKNER-----YYCVLPNDHSHSVHYIYSIQRALDCELAEE--A 256
Db 217 ESILQDEKTSRLSENKYGIDIDSCNNYSVLWMDHSHKSFQFYEQITATL--ELPNNVFG 274
QY 257 QLTHTTAIDREGRRAVAGAAACOEAKEDIK--SHSENVSOHPLHVEVLHSEIMAHOKF 313
Db 275 KMANIINDIGR-----ACIVETNLIKELLKIGOKLAQINLAVSIRSMRDIPEES 325
QY 314 ALRGLSMNMKIMYS-----SDRFQIFCOACLR----- 341
Db 326 CAVLLEWLADIAGSGICGRNRYFSSVICKELVPRWNCGLHNSDLTFLRLSLRSLALPEIVA 385
QY 342 -REPD-----SENPLCLSRM----- 356
Db 386 IDSPDIFLNEDHINSSGPDSTSHMETDESSHSRHWYPSNSLPDVLVSASVREDYFF 445
QY 357 LMDAKLYGARKILHELIFSSPEMEXKLFAMEFVKYKOLQKEYISDDHDSISITA 416
Db 446 LYDLKLWKSRLYKQLQELYGYFITQPGFKEIMGARTAISYRRLAEUFLLLDREPEHSVIF 505
QY 417 LSVQMETPTLARHLEEQNVISVITETLEVL-----PEYLD-----RNNKFNFOGY 464
Db 506 FSNQIFTVADVAKLLVTEYDELTTINATLYTFTYKKNLTPNYVDQHAMIRTSAAFH-- 563
QY 465 SQDKGRVAVTCDLKYILISPTIWTIERLR--MQLEGRSFLKILTCMOGMEETIRQV 522
Db 564 -----SRRIYHIFHTIQFMSIPCV-AEIVREDLKLQYADPFNL--PQGMCPYTRAV 614
QY 523 GQHIEVDPO-WEAAIAIQOL----KNILLMFOEM-----CACDELLVAVKECHKAV 571
Db 615 SQHVENDSMWVVLVSVLOAKLKHGVGNFMELNTNKLANINLISILYLPKARNES 674
QY 572 MRCSTFISSTKTVQSCGHSLETSKYRVSDELVSITHLPSRLTAGLHVLRSLRGAVSRL 631
Db 675 W-TNTESLTGITVDER-GNS-KLIEYDALQPVSEHPLHLLVYL-----L 719
QY 632 HEFVSPEDE-----QVEVL--VEYPLRCLVLVAQVVAEMRRNGLSLISOVFYQDVKCR 684
Db 720 SFYVERDNTKLLMTQLDILLAVTDHPLRVCAWLSQMRKALWIRNGTTLRQDAHRYNLSFH 779
QY 685 EEMYKDIIMLOIGASLMPNPKFLLVLQRYELAEAF--NKTISTKDDOLIKOYNTLIEE 742
Db 780 EYTFDLDVLLQLTLTYGPDAILPSFISRFQLEDQMYGFFVPHHYD-VSQVTIMMEE 838
QY 743 MLOQLYIYIGERYVPGVGNVTKKEVTMREIHLILCTPEMPSHAIAKNLPENNETGLEN 802
Db 839 FLLLLISVCNTAVLDHWDITRR--IEYGHAILCPRLPYSEITKRTCEHLLHKKOFES 896
QY 803 VINKVATFKK-PGVSGHYVELKDESLKDFNMVYFYHYSKTHQSKAEHMOKKR-RKOENKD 860
Db 897 TLKKVATPRNAEGINDSGSFTLKDEYFDVYDPNFIHYSRNQREAEENILRRYSKQHSKH 956
QY 861 EALPPPPPEFCFAPSK--VINLLNCIDIMYIL-RIVFERAIDTDSNLWTEGLMQNAFHI 917
Db 957 --LESVYVEYHPILHSNITIPILQSDSFVGLWHHTIVAYIYIPYDQGLGLVNTALHA 1014
QY 918 LALGLEEKQOLQKAPKEEVTFDFYHKASRGSSAMNOMLLEKLGIPOLEGOKD---- 973
Db 1015 CLLVLMSK-----GSEPIFSKKICE--NRFPVVEGLQEVCSN 1050
QY 974 ----MITWILQM-----FDTVKRLREKSLIV-----ATTSGESIKNDE 1009
Db 1051 PDVTLFSLVLCQMKHNRNFVYVK--EKISLIMKILASEVPLLYEPVYAEFTLSISSKIVQ 1107
QY 1010 ITHDKKAEKR-KKAAARLHROKIMQAQNSALQKNFIETHKLMDYNTSEMPGKEDSIMEE 1068
Db 1108 SLSDABQEQOHLAKVAKRMAREQRIEMEQFRMQONKPELENHALFEASDCEN----DEADEP 1163
QY 1069 ESTPAVDSYRSLAGPKRPSVTEKEVLTCILCOBEQEVKIENNAMVLSACVOKSTALT- 1127
Db 1164 SVTSSVS--TKFLDLP-----PIDTCLLCOBEELDKRPGYGTTLV--VLRSSVRLR 1209

QY 1128 -----QHRGKPIELSG-----EALDPL----- 1144
Db 1210 FPADANNVSEVLDIPDSDLHDEIQERPFGLAGKRKKVLDDSTAYDYDNYEYKKGKELHQ 1269
QY 1145 -----FMDPDLAYTGTGCGHVHMAVCWOKYFEAVOLSSO---ORTHVDLFDLESGE 1194
Db 1270 LKDSFENGPPQOLDRLGHLATGCGHFHHDICFNHHTATVTLATRANPYRNHP--HNLSMKE 1327
QY 1195 YLCLPCKSLCNTVPIPIIPQPKINSENADALAQLLTTLARWI-----QTVL-- 1240
Db 1328 FLCPLCKALCNTIPILWRPKKEEINFQEGAVLT--APLKNLWVSKTFSFNKDLNQQLDDI 1385
QY 1241 -----ARISGVNIRHAKGENPIFFNQMGMDSTLEFHSILSFGVSESSKYSNSIKEMV- 1294
Db 1386 ETPSEHTQSYNL-----NLLDV-LQHTLRDSLKDIYTLNTGADNS---SDNVEENAD 1434
QY 1295 ILFATTIYR-----IGLKVPDPDERPRVPMCLTWSTCAF-----TIOALENLGLDE 1339
Db 1435 NLFOSSVLDHWHFKSVVNEVPADER-----LAISDDIFELYRRLDDVIDLNSLSYSD 1488
QY 1340 GKPLFGLQNRQHNGLKALMQFAVAQ-----RITCFQVLI-----QKHLVRL 1382
Db 1489 FIPVNGKL---HNWVK-LFSYSLCQVEASTRGHIKCSSIPADIMVHNLGKQOVFLRL 1543
QY 1383 SVVLPNKSDETPCLLSIDLFHVLVGAVALP-----PSLYWDDPV---- 1422
Db 1544 S---ESIKTYTLLC--AHDQSKRIGGSIQEFEFISFCQOKRIFGRLLPSL--DSPVTKSI 1596
QY 1423 -----DLQPSVVS-----SSYNHL-YLFHLITMAHMLQIILLTVDLGLPL 1460
Db 1597 TDDRVEPLLKDTREFAEASVSGLLSCDESFHYLTQLYTADIVRNLMILLSQRNSL-L 1655
QY 1461 AQOVEDSEEA-----HSASSFFAEISQYTSIGSGICDI-----PG--WYLMWS 1500
Db 1656 KCMESVEFEADYEQLGKGFELHVIQIWSLRVDGAGLNFDCCTEDDLNPNHLLTLYKL 1715
QY 1501 LKNGITPYLRCAALFFHYLLGVTPPEELHTNSABEYSA LCSYLSLPTNLF---LFQBY 1557
Db 1716 LERFSLIFLRCALLMYCRYGVSFETPNLNFQNSLSRLQTMHPIGVIELSNHLCFLA 1775
QY 1558 WDTVRPLLRWCADPALLNCLKOKNTVVRPEKR---NSLIELPDDYSCLLNOASHFR 1613
Db 1776 SSTWLSLTKHMC-----NFTETGCLCDFPRAYYPGYVELSLPYELDKVFELLAR 1829
QY 1614 PRSADDERKHVPLFCGAILCSQNICQEIYVNGEVEGACIFHALHCGAGVCFIKIR 1673
Db 1830 SKCL-TEPMEPAICLFCGKLLCFQSHCCS--FNG--IGECNLHMQCCASDIDGFLIVK 1884
QY 1674 RVVLVEGKARCAYPAPYLDYGETDPLGKRGNPJHLRSERY-RKLHLVMOQHCHIEETA 1732
Db 1885 AILYLNPVPV-GSFVAPFLDAYGETDLGLRRGRSOYLSQKRYDETVRTMNLNGSIPSYA 1943
QY 1733 RSQETNOMLFCFNQOLL 1749
Db 1944 ROLDANPDGTG--WETL 1958

RESULT 4

T37711

probable n-end-recognizing protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C:Accession: T37711

R:Murphy, L.; Harris, D.; Wood, V.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, June 1997

A:Reference number: 221738

A:Accession: T37711

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2052 <NR>

A:Cross-references: EMBL:297208; PIDN:CA010108.1; GSPDB:GN000666; SPDB:SPAC15A10.11

A:Experimental source: strain 972h-; cosmid c15A10

C:Genetics:

A:Gene: SPDB:SPAC15A10.11
A:Map position: 1

Query Match 9.9% Score 914: DB 2: Length 2052;
Best Local Similarity 20.4%; Pred. No. 2.3e-48;
Matches 447; Conservative 326; Mismatches 734; Indels 682; Gaps 89;

QY 62 QESVQMSIETPLEW-YLFGDPP-----DICLEKLKHS-----GAFOLCGRVF 103
DB 43 QEVFLSLLNEDNWKYFLKEKPGAITSDPRLSLRLOHSEPECAQELQDKRSKGVCHGVF 102
QY 104 KSGETTSYSCRDCAIDPTCVLCMDCFQDSVHKHRYKMTSTG-GGFCDCGDEAWKTCPE 162
DB 103 RAGEVIVRCKNCGLDNVCILCAPCFHATNHEGHETHSVISSTSYSGICDCGDEAWNVN 162
QY 163 CVNHEPGRAGTIKENSRCPLNEEVIVQARKIFPSVIKVVEMTW----- 207
DB 163 CKIH-----NVPDEQKXPEVI-----PLELOHSITTTIILLDFILDFVFCSPV 209
QY 208 -----EEKELPELQIREKNERY-----YCVLFENDEHSHSDHYIS 244
DB 210 NLKASTVGSILADEAS-----RLSSAKYGVADRPCNFRVLMWNVDEVHTFADAVGS 262
QY 245 LORALDCELAEAQIHTT-AIDKEGRRVAKAGA----- 275
DB 263 VLEALDSSNTAFGLEVAQRVDSIGRFVATVSASVHEAIRIANAIKENLAVNVRTARDF 322
QY 276 -----YAA-----CQE----- 281
DB 323 REDICGILLEMFDOLLESHVCYFADYLIQIIVCDEILKNWSPGLEKPAKPEYFNFLPLEI 382
QY 282 -----AKEDIKSHSENS-----OHPL----- 298
DB 383 VNDSDSEDDIYAAEELLDVIANLODETGVTRIANGDGEDEFADMTDPTTAGFDHPLDD 442
QY 299 -----HVEVLHSEIMAHQF-----ALRLGSMNKMYS----- 328
DB 443 NDVNLLDFETERIDIDLTDVEMTEENEAADYPCVNRNTRQDDVDQDISMETESONE 502
QY 329 -----SDFRQIF----- 335
DB 503 TDESONTENVDPQHTPVPITPTATQDVVTVIRPEFNSQLLNLRQIINARRPRPAAV 562
QY 336 COACLRRE-----PDS-----ENCLJSLR-----MLWDAKLYKGARKILHELIFSFF 379
DB 563 CQVSLSDYKWSHPPIPPSSYFVSFSSSLRDYFDLDFLKWRLRGLLSKLYVVPFN 622
QY 380 MEMEYKILFAMEFYKYLQKQKEYISDHDHRSISITALSQVMTVPTLARHLIEEQNVIS 439
DB 623 RNLFLKRLMGIIRFVIHYRSLATAFADRPDHDHSMFLSVQFTTSLAEAVVKDYDFLT 682
QY 440 VITETLLEVL-----PEYLORNKNFNGYSQD-----KLGKVYAVICDLK 480
DB 683 NLNATLTLSSLTQSNRPSTL-----FSSDIETPTIQLNRQVLKTRTRYNLFSDLG 732
QY 481 YTL-----LSKPTIWTIRLMQFLBGFBSFLKILTCMOGMBEIRQVGHIEVDPDWAAI 536
DB 733 YLQHPQVKLLVDVDRYVHOYIDLLRV-----QGVIPQQRAILSHVO-----WDFPH 781
QY 537 ATOMQLKNILMFQEWACDEEL-----LLVAYKECHKAVMRCSTSFSSSKTVVQSCGHS 592
DB 782 G-----KNILFVMQVAMLSNTVSSCFQAPYERLFAI-KCIITSITHPKLDI-----AES 832
QY 593 LETKS-----YRVEDLVSIHLPLSLRTLAGLHVRLSLRGAVSRLEHFSV 636
DB 833 LEPLISCPSSSLNFTQPLVPFVSVPDPSIYHP-----LHMWLSNLFYSYCRVDASSH 885
QY 637 FEDQVQVILVEYPLRCLVLVAQVVAEMWRNGLSLISQVFFYQDVCKREEMWKDITMLQ 696
DB 886 WDKDTLLALLDHPRLVRCVLLAQIDCNLWIRNGRSILLTDAFYLRNNIEVSYDKDILAI 945
QY 697 IGASLMDPNKFLLLVLQRYELAE-AFNKTTISTKDQDLIKQVNTLIEEMQLVLIYIGERY 755

DB 946 TILMEVDPNLVNAVQVQFEFTDWLYNLTYNEHPNYDTERIPAMLCMKLELLIALITER- 1004
QY 756 VPGVNVTKEEVYTMREIHLICIEPMHPASHIAKKNLPENENNETGLENVINKVATFKKP-G 814
DB 1005 -EQIHLVDIQQIIRTLAQOLCGPLAYSALLSTISSNLVESLSPKIREVEVTSYKAPDG 1063
QY 815 VSGHGVYELKDESLKDFNMFYHYKSTQHSKAFHMKKRRKQEN--KDEALPPPPPEFC 872
DB 1064 LHDFGVYSLKDEYDLVDPYFYHKNKNERESDTILKRLAKKNVSAESIIEPKIRFL 1123
QY 873 -----PATSKVINLNCIDIMMYILRTVFRADTDNLMWTEGLQM-AFHILALG--LLE 924
DB 1124 EKGDHDIFFAAVNASTFSLIIF-----RATE-----YALVQAESFGSDIGNITLG 1169
QY 925 EKQQL-----OKAPEBEVTFDYHKA-----SRLGSSAMNIQMLELKGIPQL 968
DB 1170 DALQJLCLSMKTHEFSKSNDFCSRCAERYPTDSSIMREFGGSAYCLAEJCFAILKSPKY 1229
QY 969 EGOKDMITWI---LQWFO--TVKRLREKSLIVATTSGSESINKDEITHDKAEKRRKA 1023
DB 1230 KDVHVKNVAVLAGLQKNPDSAYSNMLEATHFELSTTSS-----TSDSNEIEKTQEK 1280
QY 1024 EAARLHRO-KIMAQHSALQKNFIETHKLMYDNTSEMPGK---EDSIMEEESTPAVSYSR 1079
DB 1281 KRLALEKOKKIMQFRDQOASFIA-----QNTDFDIGEDQTEDEVTTTEPEEVKYHEH 1334
QY 1080 TALGPKRGPSTVEKVLTCILCOEQEVEK-----IENNAMVLSACVOKSTALTQHRG 1131
DB 1335 I-----RG-----NCLLCOEENQDQAPYGVIGIIQSSLLRKTVDHSEILDEIYS 1380
QY 1132 KPIELS-----GEALDPLFMDPD-----LAY-----GTYTSCGHVMHVAWCOKY 1171
DB 1381 VPPNLDRESHSPFGKKYDTVVFNRSKDRLLSAYPCGNIRGVFVSCGHLMLHLCFKNY 1440
QY 1172 FEAVOLSSQORIHVLDLDES-G-YY-----LCPLCKSLONTVPIIPILOPKIN 1219
DB 1441 YVARSWYRN-----DVTAGLSEYKYKSTAKFEMCPCLCRSLNVLLP-MFOIPKM-- 1489
QY 1220 SENADALAOQLTLARWIOQTVLARISQYNIHAKGNPIPIFFNOGMGDSLEFHSI---- 1275
DB 1490 CLNIDTLNFPFRMNGWLEEI-GTMSSSSFEYQ-----LVRSSLSDDKDTFRSCFLRP 1540
QY 1276 -----LSFGVESSIKYSNS--IKEMVILFATTIYRIGLKVPPDRDPRV-PM 1319
DB 1541 WINSKLIISAMLARLKIADGALLDQSNRNVSDLYDRYCETT-KLAMKLVKSGFTTNNVSPH 1599
QY 1320 LTWSTCAFTTQAIENLLGDEGRP-LFGALQNRQHNGLKAL-MQFAVAQRITCQVLIQKH 1377
DB 1600 DLLNSLAYTVSSLE--VSQRCSPKQSGATRSVWFMELGPLTLFLPLTSDT----- 1648
QY 1378 LVRLLSVLPNLIKSEDTPCLLSIDLPHVLGAVLAPPSLY-----WDDPVDLQPSV 1429
DB 1649 ---VLKVCVDQIISKDQOALLMESOKLLVCKIFYRHSOLKSLMRNGRMSDHDQIOFLL 1705
QY 1430 SSSY-----NHLXFLHITMAHMLQILLTVDGTGLPLAQOVED---SEE 1469
DB 1706 STNFDQFVKISSMLIFGQDNILYVVKLFYLSCKTIISMIKKVADVSVVPLDTINYS 1765
QY 1470 AHSASSFAAEISQYTSGSIGCDIPGWYLVWSLKNGI-----TPYL 1509
DB 1766 QQSKSQFY-----ILC--KNVLLCGSSNNIEILDDESLLRLMSLVEKYSPLFL 1813
QY 1510 RCAAALFFHYLLGVTPP-BELHTNSAGEYSALCSYLSLPTNLLFLLQEYW-----DT 1560
DB 1814 RRVALVLYCMFDISLEFNEFSNNEDESELERLSKLKIVPP---LQELYSQMSDENNQ 1868
QY 1561 VRPLQRACADPALLNCLKQKNT-----VVRYPRKRNSLIE-LPDDYSCL 1604
DB 1869 ILELAGWCEHLA-----QNTWGDSTISLEYPGIYELVKPLHRLNLLDSMOESVCCM 1921
QY 1605 LNAQSHFRCPRSADDERKHPVLCFLCGAILC-----SONICQOEIVNGEVGACIFHALHCG 1661

Qy	1152	YG -TYTSGCG-----HVMHVCWKQYFEAVOLSSQQRHHVDLFDLESGEYCLPCK	1201
Db	1278	EALKENGSCGRKVFVSCNHHTHHNCFKRYVQKKRFS-----SNAFICPLCQ	1324
Qy	1202	SLCNTVPIPIPOKQINSENADALAOQLTLTARWLTQTVLARISGYNIRHAKGENPIPIFF	1361
Db	1325	TFNSCTLPL-COTSKANTG-----LSLDMFLESELS-	1355
Qy	1262	NOGMGDSLTFFHSILSFGVSESSIKYSNIKEMWILFATTIVIRIGLKVPPDERPRVPMLT	1321
Db	1356	-----DTLS-RLPKPFTTEENRPTNIFSIMI-----	1381
Qy	1322	WSTCAFTIQAENLLGDECKPLFGALQNRHQHGLKALMQFAVQRITCPOVLI-----Q	1375
Db	1382	-SOCQFDDKAVRKANFRSHKDVSLLSVHMANTISML--EIASLEKPYISIFFRSREQ	1437
Qy	1376	KHLVRLLSVVVLNPKSEDTPCLLSIDLPHVLGVAVLPPSLYWDDBDVLDLPSSVSSSYNH	1435
Db	1438	KY-----KTLKNI-----LVCIMLETFVIGK-----PSMEF-EPYPQPDVTWNO-NQ	1478
Qy	1436	LY-----LFHLITMAHMILOILLTVDTGCLPLAQVOEDSEEAHSASSFFAEISQ-----	1482
Db	1479	LFQYIVRSALFSPVSLRQTIVTEALFTFSRQFLRDFLOGLSDAEQVTKLYAKASIGDVLK	1538
Qy	1483	-----YTSGSGCIDPGWYLVWLSKNGITPYLRCAALF---PHYLL	1520
Db	1539	VSEQMLFALRTISDVMEGLDSESIYDLAYTFLLKSL---LPTIRCLVFIKVLHELIV	1594
Qy	1521	GVTPPEELHTNSAEGEYSALCSYLSLPTNLFLLFOEYWDTVRPLLQRCWADPALLNCLKQ	1580
Db	1595	KDSENETLVINGHEVEE-----LEFEDTAEFVNKALKMITESLESVLDLTT	1641
Qy	1581	KNIVVRYPKRN-----SLIELPDDYSCLLNQASHFRCPRSAD-----DERKHPV	1625
Db	1642	QESIVSHPYLENIPIEYCGIITKLIDLSKYLTNTYQTSKEIKLREERSQHKMKNDRLDFK	1701
Qy	1626	LCLFCGAIL-----CSONICQCEIVNGEEVGACIFHALHCGAGVCIFLK	1669
Db	1702	ICLTGCVKVLADRHMTKHLNKCFFPGFGLMPNSSE-----VCLUHLT	1747
Qy	1670	IRECRVVLVEGKARGCAYAPYLDVEYGETD-PGLKRGNPLHSRERYKRLHLVMOHCII	1728
Db	1748	QPPSNIFI-----SAPYLSNHGVEGVRNMRGDLTTTLNKLRYEHLNRLWINNEIP	1797
Qy	1729	BEIARSQE-----TNQMLFGFN	1745
Db	1798	GYTSRYMGDEFRVTILSNGFLLFAFN	1822

RESULT 6

T30554

ubiquitin-protein ligase E3 - yeast (Kluyveromyces marxianus var. lactis)
C:Species: Kluyveromyces marxianus var. lactis, Candida sphaerica
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jan-2000
C:Accession: T30554
R:Waller, P.R.H.; Varshavsky, A.
submitted to the EMBL Data Library, April 1998
A:Description: Kluyveromyces lactis UBRL, the recognition component of the N-end rule
A:Reference number: Z20856
A:Accession: T30554
A:Status: preliminary; translated from GB/EMBL/DDBB
A:Molecule type: DNA
A:Residues: 1-1941 <NALL>
A:Cross-references: EMBL:AF061554; NID:g3114670; PID:g3114671; PIDN:AAC15841.1
C:Genetics:
A:Gene: UBRL
C:Superfamily: ubiquitin--protein ligase

Query Match 7.0%; Score 647.5; DB 2; Length 1941;
Best Local Similarity 20.5%; Pred. No. 8.9e-32;
Matches 409; Conservative 319; Mismatches 689; Indels 581; Gaps 91;

Qy	35	FYTAFLHHLAQLVPEIYFAEDMDPLDKBQESVQMSFTPLLEWYLFCEDDPDICLEKLUHSG	1
Db	51	YYYVIMSDSGRLLPHMFTATNEREFKPNVDQAMEIKLSSKPWYKIDENG--H\$KENHAG	108
Qy	95	AFQLCGRVFSGETTYSRCDAIDPTCVLCMDCFQDSVHKHRY--KMITSTGGGFCDCG	152
Db	109	--RICGAKFRVGEPIYCKECSDDTCVLCVNCNFPKDHVGHVYTSICTEFNNGICDGG	166
Qy	153	DTEAWKTPGCVNHEPGRAGCTIKENS-----CPLNEEVIQARKTFPSPVIKY	200
Db	167	DKEAW-----NHELCKGA-EDNGLEDFDDHDKGK\$KML\$ESYLIELFHDHFDVFNQ	218
Qy	201	VVE--MTIWE-----EEKELPPELQIREK-----NER-----	225
Db	219	NIEPLTTIQRLPLAKLRYFNPEREIYEOADMLRRLAYRNQYMDDESSNKRHLTSLDPLS	278
Qy	226	---YVCVFNDEHHSYDVHVIYLSQRLADCELAELAHHTTAIDKEGRRAVKAGA-YAACQ	280
Db	279	TLKDYAILVYVYDFHNYSQASAAIROG-GPDNKHIDLLTAKIDSEGR\$LLRCSADIASLM	337
Qy	281	EAKEDIKSH--SENVSOHPLHVEVLHSEIMAHOKFALRLG\$WNKIMSY--SSDFROIFCQ	337
Db	338	GRIFSVO\$NGL\$CTITO--WYELHQEAC---KY\$T---MWINDCLNPN\$T\$QSLFRN	388
Qy	338	A-----CLR\$EPD-----SENPLCI-----	352
Db	389	AIGVK\$K\$YEPFQ\$IDMT\$VVDYFSD\$YLSDDPYLADH\$VLGEGYKPIPLGRH\$KSLD	448
Qy	353	-----SRU---MLMDAKLYKGARKILHELIF\$SF\$FMEYK	385
Db	449	PGDISA\$PI\$LNK\$VIAEDH\$HEY\$N\$R\$Q\$Y\$V\$FL\$EN\$Y\$W\$K\$K\$R\$K\$IV\$Q\$D\$LI\$PT\$	508
Qy	386	KLFAME\$V\$Y\$K\$OLO\$E\$Y\$I\$D\$D\$H\$R\$S\$I\$T\$A\$L\$S---V\$M\$E\$T\$V\$T\$P\$L\$A\$R\$H\$	435
Db	509	PMF\$D\$Q\$V\$E\$I\$F\$P\$H\$M\$R\$S---G\$T\$F\$M\$D\$R\$P\$Q\$U\$T\$S\$R\$E\$V\$V\$Q\$F\$T\$	566
Qy	436	NVISVITE--TLLE--VLPEYLD\$RN\$K\$F\$N\$F\$Q\$Y\$S\$Q\$K\$L\$G\$R\$V\$A\$V\$IC\$D\$K\$Y\$	487
Db	567	SVIDV\$F\$D\$T\$M\$D\$E\$G\$L\$V\$W\$O\$R\$V\$O\$R\$N\$P--\$K\$S\$Y\$S\$I\$F\$K\$Q\$G\$Y\$A\$V\$E\$T\$	624
Qy	488	TIWTERLMOFLEGR\$F\$R\$K\$ILT---CMOGMEET--RQV\$G\$HI--EVD\$P\$W\$E\$A\$A\$I\$O\$M\$Q\$	542
Db	625	-----GEFIMIVT\$L\$CK\$L\$F\$N\$G\$A\$W\$K\$K\$K\$R\$E\$G\$H\$V\$L\$R\$E\$Q\$H\$F\$P\$Y\$E\$V\$T\$S\$V	669
Qy	543	KNILLMF-----Q\$W\$C\$A\$D\$E\$LL\$V\$A\$K\$E\$C\$H\$A\$V\$M\$R\$C\$T\$S\$F\$I\$S\$S\$K\$T\$V\$V\$Q\$S\$G\$H\$	591
Db	670	YSIIQT\$D\$K\$V\$Q\$O\$S\$K\$H\$ID\$Q\$R\$L\$IGA-----INLL\$D\$F\$L-----G\$H\$R\$N\$S\$Y\$K	711
Qy	592	---SLETK\$Y\$R\$V\$E\$D\$L\$Y\$S\$I\$H\$L\$R\$T\$L\$A\$G\$L--H\$V\$R\$L\$S--R\$L\$G\$A\$V\$R\$H\$E\$F\$Y\$S\$F\$E\$D\$F\$Q\$E\$V\$L	645
Db	712	LYKDFEIIK\$Q\$I\$S\$K\$E\$Q\$Y\$F\$M\$N\$P\$H\$T\$F\$C\$F\$L\$V\$H\$V\$P\$LO\$V\$S\$I\$Q\$V\$S\$K\$D\$Y\$L\$V\$S\$D\$F\$	766
Qy	646	VEYPLRCLVLVAQVVAEM\$R\$R\$N\$G\$L\$S\$L\$S\$Q\$Y\$Y\$O\$D\$V\$K\$C\$R\$E\$M--Y\$K\$D\$I\$M\$L\$O\$I\$G\$A\$S\$L\$M\$	703
Db	767	---AL\$R\$V\$L\$C\$S\$Q\$ID\$G\$F\$W\$R\$N\$G\$M\$V\$L\$H\$O\$S\$A\$Y\$---K\$N\$N\$P\$E\$M\$S\$Y\$S\$R\$D\$I\$Q\$N\$L\$Q\$--A\$F\$L\$I\$E	819
Qy	704	PNKFLLV---L\$Q\$Y\$E\$L\$A\$E\$A\$F\$N\$K\$I\$T\$K\$D\$O\$D\$L\$K\$Q\$Y\$N\$T\$L\$E\$M\$Q\$V\$L\$Y\$I\$V\$G\$R\$Y\$P\$G\$V\$G	760
Db	820	KND\$Q\$R\$Y\$V\$N\$N\$L\$D\$R\$W\$E\$L\$D\$W\$F\$D\$G\$S\$V\$P\$T\$E\$T\$V\$Y\$D\$K\$I\$S\$S\$I\$Q\$Q\$E\$V\$A\$F\$L\$Q\$Y\$L\$V\$R\$	873
Qy	761	NVTKEEY\$T\$R\$E-----I\$H\$L\$L\$C\$I\$E\$P\$H\$S\$A\$I\$K\$N\$P\$E\$N--E\$N\$N\$T\$G\$L\$E\$N\$V\$K\$N\$V\$A\$T\$F	810
Db	874	DFYK\$K\$F\$D\$T\$L\$E\$T\$Q\$Y\$N\$T\$K\$N\$A\$I\$Y\$K\$L\$A\$E\$P\$L\$S\$Y\$T\$D\$L\$N\$D\$I\$D\$Y\$L\$T\$E\$S\$Q\$F\$D\$T\$V\$E\$E\$S\$T\$Y	933
Qy	811	K\$P\$G\$V\$G\$H\$Y\$E\$L\$K\$D\$E\$S\$L\$K\$D-----F\$N\$M--Y\$F\$Y\$H\$Y\$S\$K\$T\$Q\$H\$S\$A\$E\$H\$M\$K\$R\$R\$K\$O\$E\$N\$K\$D\$E\$A	862
Db	934	LEP\$K\$S\$I\$W\$C\$F\$Q\$T\$K\$E\$G\$T\$Q\$E\$N\$R\$F\$L\$N\$M\$G\$N\$D\$E\$H\$S\$A\$T\$I\$V\$K\$S\$H\$L\$A\$D\$S\$K\$E\$K\$R\$A\$I\$V\$K\$P\$Q\$	992
Qy	863	L\$P\$P\$P\$P\$F\$C\$P\$A\$S\$K\$V\$N\$L\$N\$C\$M\$D\$I\$M\$Y\$I\$L\$T\$V\$T\$E\$R\$A\$I\$D\$T\$S\$N\$T\$W\$T\$E\$G\$M\$O\$M\$A\$H\$I\$A\$L\$G\$L	922
Db	993	-----L\$E\$L\$D\$E\$L\$D\$P\$C\$A\$R--E\$L\$G\$S\$F\$T\$R\$T\$N\$F\$A-----K\$L\$I\$F\$K\$L\$K\$L\$A\$V	102

[illegible]

Qy	245	LORALCELAELAQHLHTTAIDKEGRRAVKAGAYAACQAKEDIKSH-----SENVSQ-H 296
Db	266	MENAIQIEEEECNVHYMDLASTITRIUNTEPEVAISITRALEDSDHVVTVLQSENFEID 325
Qy	297	PLHVEVLHSEIMAH-----QKFAIRLUGSNMNI 324
Db	326	QIAKEFOKENIVHVRKADDIFKRKLTDLTDWLYSLCFAATSLQNKYALRI-SMLDVM 384
Qy	325	MSYSDFRQIFCQACLRPEPDSNP-----CLISR----- 354
Db	385	YSHESK-----MRVSPNTNPDFSKNLLGGFLISNEDSWEFKPWSLENIEDER 435
Qy	355	-----LMLWDAKL-----YKGAR-----KILHELIFSSF 378
Db	436	ISKTLTNYNRLITRAHSPNTVSHFYNYGSRFOYIININILSKSKFKML-KIMASUL 494
Qy	379	FMEMEYKFLAMEFVKYKQLOKEYISDDHRSISITALS-----QMFTVPTLARHLIE-- 433
Db	495	SLRDESRKFLAAQYIDVLSVLYDAVASD-AKEQVTLMSILQGYTFQDPSIANMTISSG 553
Qy	434	--EONWTSVIT-----ETLEVLPEYLDNRNNKFNFOGYS--ODKIGRVYAVIC--DLKY 481
Db	554	FIERTIRFAETLMAFNPEDLMSVLPISL-----YNGKLPETITRNRRITICFKDLCT 606
Qy	482	IL-----ISKPTINTIERUMQFLEGFRSFLKILTCMQGMEIIRQVGOHLEVDP-DWEAAI 536
Db	607	IMSAWTPHEELLSNEAIFNAIESFSFSNVL-----PLKRETKEHVEVENDFS--- 656
Qy	537	AIOMOLKNILLMFOEW-----CACDEBELL-----VAYKECHKRAVM--RC 574
Db	657	AFYEFFSSILIMTDGYTRISLVKDAAFKROKVLKLDVAOTREFFSLTNSRKAIASPDNA 716

RESULT 8
S64851
probable membrane protein YLR024c - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein L1730
C;Species: Saccharomyces cerevisiae
C;Date: 01-Aug-1995 #sequence_revision 24-May-1996 #text_change 05-Nov-1999
C;Accession: S64851

R;Obermaier, B.; Piravandi, E.; Rinke, M.
submitted to the Protein Sequence Database, May 1996
A;Reference number: S64845
A;Accession: S64851
A;Molecule type: DNA
A;Residues: 1-1872 <OBE>
A;Cross-references: EMBL:Z73196; NID:g1360331; PID:e245500; PID:g1360332; G
A;Experimental source: strain S288C

```

C;Keywords: transmembrane protein
F:199-205/Domain: transmembrane #status predicted <TM1>
F:804-820/Domain: transmembrane #status predicted <TM2>
F:899-905/Domain: transmembrane #status predicted <TM3>
A;map position: 12K

```

Query Match 6.0%; Score 549.5; DB 2; Length 1872;
Best Local Similarity 19.8%; pred. No. 1.1e-25;
Matches 388; Conservative 321; Mismatches 729; Indels 521; Gaps 94;

[illegible]

```
QY 1114 -----MVLSACVOKSTALTOHGRKPIELSGEALDPLFMDPLDAYTYTGSG 1160
Db 1270 IDFTNPTDVNRINSLSFGKQKSAIQEN---PQDDGTRLFTKTCPEVLR-----ACG 1320
QY 1161 HVMHAVCWQKFEAVQ-LSSQQRHVDLFDLESGEYLCPCKSLCNTVIP-----1209
Db 1321 HGSHTKCLSGHMKSTIRGIONOTTKNIPL-SYSGGLIYCPVCNSLSNPLPKTNDIDKRTS 1379
QY 1210 ---IIPLOQKINSNADALQALLTLARWQTVLARISGYNIRHAKENPIPIFNQOGM 1266
Db 1380 SQFFMCIEKRSEAEENLPMSSICAKAMI-----LG 1411
QY 1267 DSTLEFHSILSGVESSIKYSNIKEMVLLPATTIYRIGLKVPPDERDPRVPM-L--TWST 1324
Db 1412 D--LOGKKVTT--IEDAYKVNSV-----FINTISNLRSHKHKGIKVMNERISSQ 1461
QY 1325 CAFTTQATENILGDGKPLFGALQNRHNGLKALMQFAVAQRI--TCPQVLIOKHLVRLLS 1383
Db 1462 CILTLHLVCELKSFYKKFVNS-----KTESSELSRKINWNEFLIKGNVNL 1511
QY 1384 VLPNI-----KSEDTPCLLSIDLH-----VLGAVLAFPSLYWDDVDLPSPSVS 1430
Db 1512 YMSQNFNDIGCKTPOPPNLCIYEMFKRRFHQLLLLLARDMMRVNFYKDCRNKIKISSNG 1571
QY 1431 S---SYNHLIYFLHITMAHMLQILLTVDTGLPLAQVQDSEAHASASSFAELISQYTS 1487
Db 1572 SEEPSTSESYLFN--TFKKYVDLFKPPD-----VRFDTSLKIKDFICSL-----1615
QY 1488 IGCDFPGWLVSLKNGTTPYLRCALFFHYLLGVTPPEELHNSA-----EGEYSAL 1540
Db 1616 -----LESUSIFCRFTFLFNQYDDGDDGNNNSNMDVMKQREIELI 1662
QY 1541 CSYLSLPTNLFLFOEYWDTVRPLLRW-----CADPALLNCLKQKNTVVYRPRKR- 1592
Db 1663 FRYFKPLNLTFLKDFYFNELTQNIERYNDGNDNLRIQQVIYDMVQNIINTRA-YPSPEHI 1721
QY 1593 SLIELPDDYS--CLLNQASHRCPSRADDKHPVLVLCFGAILC--SONICCOEIVNGE 1648
Db 1722 QLIELPLNLSFLSNDNDSNKK-----DKYEIAVCLCGQ-KCHIQKSIALQGYLOGE 1774
QY 1649 EVGACIFHALHCGAGVC---IFLKIRECVVLVEGKARGCAYPAPYLDEYGETDPGLKR 1704
Db 1775 ---CTDH-MRNGCEITSAYGVFLMTGTNAIYLSYK- RGTFYAAPYLSKYGETNEDYKF 1828
QY 1705 GNPLHLSRERYKHLVWQOHCIEBIARSQTNOMLEF 1743
Db 1829 GTPVYLNARYANL-----ANEIVEG 1849

RESULT 9
T48252
eCERIFERUM3 (CER3) - Arabidopsis thaliana
N:Alternate names: protein Tle22.70
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T48252
R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft,
submitted to the Protein Sequence Database, March 2000
A:Reference number: 224489
A:Accession: T48252
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-795 <BEV>
A:Cross-references: EMBL:AL162874
A:Experimental source: cultivar Columbia; BAC clone Tle22
C:Genetics:
A:Map position: 5
A:Introns: 203/2; 220/3; 428/1; 488/3; 666/2; 701/2; 723/3; 749/3; 772/3
A:Note: Tle22.70

Query Match 3.7%; Score 343; DB 2; Length 795;
Best Local Similarity 23.0%; Pred. No. 2.2e-13;
A:Cross-references: EMBL:AL023796; PIDN:CAA19375.1; GSPDB:GN00067; SPDB:SPBC32F12.14;
```

```
Matches 190; Conservative 123; Mismatches 274; Indels 240; Gaps 43;
QY 1033 IMQMS-----ALQKNFIETHKLMY-----DNTSEMPGKEDSIMEESTPAVSYSRIA 1081
Db 56 LMLQLSDTTISANMIESIKARLIGNQTEKRSDDGRGKDESNNM---SLEIAMYQTVR 112
QY 1082 LGPKRGSVTEKEVLTCLCOEEQEVKTENNAMVLSACVQKSTALTQHRKPIELSG--- 1138
Db 113 -----NKIENMINQSLTRVDHQPHEAEN-----CSEKNSV-----GGPSTLOGRPP 153
QY 1139 -----EALDPLFMDPLDAYTYTGSCGHMHAVCWQKFEAVQVLSQSOORI 1183
Db 154 DIRSQTSRRPDAGSDGFHPIDCD-----GYVLSGCGHAVHQSCLERLYLSKLSKERSGRT 208
QY 1184 HVD---LFDLFSGEYLCPCKSLCNTVIPITP-----LQPK-KINSNADALQALLT 1231
Db 209 VFEGAHIVDLKKBEFLCPVCRRLANSVPECPGDLCSVSKLQDSPRTKLRRKDALQPSL- 267
QY 1232 LARWQTVLARISGYNIRHA-----KGENPIPIFNQOGMDSTLEFHSILSGVESS 1283
Db 268 ---WLSALCL-----LRSAAEVIEDGDRGKTVP-----QGDGPRR-----301
QY 1284 IKYSNLIKEMVILF-----ATTIYRIGLKVPPDERDPRVPM-LTWSTCAFTTQATENILG 1337
Db 302 -KDLKSVSKMLWDYFPKPEDKTLKRLWL--PPQS-----IYVWDTLTKYLSLISME--IG 350
QY 1338 DEG-----KPLFGALQNRHNGLKALMQFAVAQR---ITCPQVLI-QRHLVRL 1381
Db 351 TRFAKNMPLPYCIDSLYEELKTSKGTILSVLLRVQSSRTKNTIHRQRFVGMKHLAES 410
QY 1382 LSVVLPNIKSED-----TPCLLSIDL-----FHVLGAV--LAPPSLYW 1418
Db 411 ICYGVSSSSSSSIFGSEGTGSLKNIDLLMNRASDPVLAHDPPFSLMVALFCLPPFLTC 470
QY 1419 DDVDQLQSPSSVSSYNHLYLFLHITMAHMLQILLTVDTGLPLAQVQDSEAHASASSFEA 1478
Db 471 EESL-----LSLVHIFHSVSVQTVIYACACRP-----SELSEUNFGENLLN 512
QY 1479 EISOYTSGS-----IGCDIFGWYLVWSLKNIGITPYLRCAALFFHYLLGVTPPE 1526
Db 513 DISNALRESGWEYFRSNMNDLSDIKD-----TIRKYSLPFLRRCALLKWL--KSTPR 565
QY 1527 ELHTNSAEGEYSALCSYLSLPT--NLFLF--QEYWDTVRPLLRQWCA DP--ALLNC-LK 1579
Db 566 KLH-----EESOMFDLPSPDPTDNMDFIYSPQSELNHHVQLEKMFNIPPIIDILNDELL 619
QY 1580 QKNTVW-----RYPRKNS-----LIELPDDYSCLLNQASHRCRPSADDE 1620
Db 620 RSSTQIWLQHFQREYRVNRVRSKLCITPVVPFQLKLPNLQDILLQRCIKKRCV-NCTKV 678
QY 1621 RKHPVLCLFCGAILCSQ--NITCCOEIVNGEVEGACIFHALHCGAGVCIFLKIRECRVVLV 1678
Db 679 IEEPVLLCLCGS-LCSPIWSPCCRE-----SCGNHAIITCGAGTGVEFLIRR-TTILL 729
QY 1679 EGKARGCAYPAPYLDEYGETDPGLKRGNPULHSRERYRKLHLVWQO 1725
Db 730 QREARQSPWSPVLDTFGEEDIDMIRGKRLYNEERYAALTYLVGSH 776

RESULT 10
T40238
hypothetical protein SPBC32F12.14 - fission yeast (Schizosaccharomyces pombe) (fragme
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T40238
R:Moreno, S.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, June 1998
A:Reference number: 221915
A:Accession: T40238
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-271 <MOR>
A:Cross-references: EMBL:AL023796; PIDN:CAA19375.1; GSPDB:GN00067; SPDB:SPBC32F12.14;
```

A: Experimental source: strain 972h-; cosmid c32f12

C: Genetics:

A: Gene: SPBC32F12.14

A: Map position: 2

Query Match 2.4%; Score 218; DB 2; Length 271;
Best Local Similarity 26.0%; Pred. No. 2.8e-06;
Matches 60; Conservative 34; Mismatches 97; Indels 40; Gaps 6;

```
QY 60 EKQESVQMSFTTPLEWY-----LFGEDPDICL-----EKLKHSACAFQIC 99
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 37 ESAKSLNLFVSALLGVHTLWNTLLPERPTIDASFLLRRAQGHSEGEYRHGTCESKC 96
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 100 GRVFKSGETVYSCRDCAIDPPCVLCMDFQDSVHKHNRHKMHTSTG-GGFCDCGTEAWK 158
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 97 GHIFRKEGVYRCVCSVDNSALCVKCFRATSHKDHSTFTVSGAGSGCCDCGNAAWI 156
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 159 TGPCFVNHEPGRAGTIKENSRCPLNEEVIVQARKIFPSPVIKYVVMETIWEBE--KELPPE 216
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 157 GDVSKIHSHEDATISNDMIDEIPEKLENSIQTTIDCVLDFVLDVFCSPENLKKMPTL 216
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 217 LQI--REKNER-----YYCVLFNDEHSHSVHVIYSLQALD 250
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 217 ESILQDEKTRLSKNKYGDIDDSNMYSLVLWDEKHSFKQFYEQITTALE 267
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
```

RESULT 11

JC5837
364K Golgi complex-associated protein - rat
C: Species: Rattus norvegicus (Norway rat)
C: Date: 05-Mar-1998 #sequence_revision 13-Mar-1998 #text_change 20-Jun-2000
R: Accession: JC5837
R: Toki, C.; Fujiwara, T.; Sohma, M.; Hong, H.S.; Misumi, Y.; Ikehara, Y.
Cell Struct. Funct. 22, 565-577, 1997
A: Title: Identification and characterization of rat 364-kDa Golgi-associated protein re
A: Reference number: JC5837; MUID: 98093490
A: Accession: JC5837
A: Status: nucleic acid sequence not shown
A: Molecule type: mRNA
A: Residues: 1-3187 <OK>
A: Cross-references: DDBJ: D25543; NID: g516825; PIDN: BAA05026.1; PID: g516826
C: Comment: This protein plays a role in the formation and maintenance of the characteris
C: Superfamily: giantin
F: 49-549, 624-1176, 1238-1707, 1763-3114/Domain: coiled-coil leucine zipper #status predict
F: 3165-3187/Domain: membrane anchor #status predicted.<MAD>

Query Match 2.2%; Score 205.5; DB 2; Length 3187;
Best Local Similarity 20.2%; Pred. No. 0.00068;
Matches 243; Conservative 182; Mismatches 472; Indels 305; Gaps 49;

```
QY 125 MCFQDSVHKHNRKYMHTSTGGGFCDCGDTFAWKTGPFCVNHPEGRAGTIK-----ENS 178
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1057 IDLLQEEITENQATIQRFITGMDAGDGSAAVKETSV---SSPRAGGGEHWKPELEGK 1112
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 179 RCPLEEVIVQARKIFPSVI-KYVVENTIMEEKELPELPQIREKNERYCYVLFNDEHHS 237
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1113 IVDLEKEKTLQKKLQKALISRKAILKKAQKEKHLKELEKQDAYRHLOEQFDQGSKE 1172
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 238 YDHVIYSLQALDCELAEAQIHTTAIDKEGRRAVKAAGAYACQAKEDIKSHSENVQ-- 295
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1173 NENIRAPLRLOAKESTDQQLPGT----GQOEPTHGSEGLSLEGTPEPASESDLHAAQPS 1227
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 296 HPLHVEVLHSEIMAHQ-----KFALRGLGSMWNKIMSYSDFRQIFCQACLREEP 344
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1228 HPGETATLQATVSAIQDQDLKEIYEVEKELEL-----KISSTTSELTKKSEEVLLLEQ 1282
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 345 DSENPCLISRLMLWDAKYKARKILHELIFSPF-FNMEYKKLFPAMEFVYKQLQKEY 403
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1283 INEOGLRIQNLKAASAKAHAHQTEQLKQKELESSQLKADLEHLKTLQPEL-----ETLQKHV 1338
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 404 ISDDHRSISITALSVQMTVPTLARHLIEBQNVISVITETLLEVLPEYLDNRNKNFQGG 463
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
```

RESULT 12

A56539
giantin - human
N: Alternate names: macrogolin
C: Species: Homo sapiens (man)
C: Date: 19-Oct-1995 #sequence_revision 26-Jan-1996 #text_change 10-Dec-1999
C: Accession: A56539; S37536
R: Seelig, H.P.; Schranz, P.; Schroeter, H.; Wiemann, C.; Griffiths, G.; Renz, M.
Mol. Cell. Biol. 14, 2564-2576, 1994

A:Title: Molecular genetic analyses of a 376-kilodalton Golgi complex membrane protein (

A:Reference number: A56539; MUID:94187728

A:Accession: A56539

A:Molecule type: mRNA

A:Residues: 1-3259 <SEE>

A:Cross-references: EMBL:X75304; NID:g405714; PIDN:CAA53052.1; PID:g405715

C:Genetics:

A:Gene: GDB:GOLGB1; GCP; GCP371

A:Cross-references: GDB:454958

A:Map position: 3q13.31-3q13.31

C:Superfamily: giantin

C:Keywords: coiled coil; Golgi apparatus; transmembrane protein

F:3238-3254/Domain: transmembrane #status predicted <TMN>

Query Match 2.1%; Score 190.5; DB 1; Length 3259;
Best Local Similarity 18.0%; Pred. No. 0.0061;
Matches 221; Conservative 207; Mismatches 424; Indels 377; Gaps 49;

QY 164 VNHEPRAGTIKNSRCPNEEVIVQARKIPPSVIKYVVMETWEEKELPPELO--IRE 221

DB 12 VLHESGDDDDTDONMRAPDPELHQLDSMEFNNTTQEDVOERLAYAP-QLVVELKDIIRQ 70

QY 222 KNERYYCVLFNDEHSHYDVHVIYSLOALDCELAFAQLHT---TAIDK--EGRRAVKA 275

DB 71 KOVOL-----QOKDEALQERKAADNKKLKLHAKAKUTSLUNKYTEEMKAQGGTV 121

QY 276 YAAQCAQAKEDIKSHSNVSOHPLVHVEHSEIMAHOKFALRLGSMWNKIMSYSDFRQIF 335

DB 122 LPTPEQSEQLSKHDKSSFEEMEIEKIRKLOKEBELISTL----- 163

QY 336 CQACLREEDPSENPCILSLMLMDAKLYKGARKILHELIFSSFFMEMEYKFLFAMEPVKY 395

DB 164 -QAQL-TOAQAEQPAQSS-----TEME-----EFVMM 188

QY 396 YKQLOKEYISDDHRSISITALSVQMFTPTLARHLIEQN-----VISVIT 442

DB 189 KOOLQEK-----EETISTLQALSTQAEAAQVVRKDAFETQVRLHEDELQGLVT 242

QY 443 ETLEVLPEYLDNRNKNFNQYSDQKLGIVYVICDLKYLIS---KPTIWTERLRN-- 496

DB 243 QADVETEMQOKLRVLQKLEHEESLVGR--AQVVDLLOELTAEOBNILSQOLOOME 300

QY 497 -----QFLEGRSFLKILTCMQGMEIRROVQGHIEVDPDWEAATAIQMLKNILLMF 549

DB 301 AEHTLRNVTETERESKILLEMELEVAERKLSFH-----NLQEEHMHLEQF 349

QY 550 QEWACDEEL--LLVAYKECHKAVMCSTSF1--SSSKT--VVQSCGHSLETKSYRVSEDL 604

DB 350 EQAGQAQAELESRYSALEQKHKAEMEERTSHLSLQKTOELQASACDALKDQNSKLQDK 409

QY 605 VSHILPLSRLTAGLHVRLSLGA----VSRL----HEF--VSFEEDQVEVLVEYPLRC 652

DB 410 NEQAVASQATIQOLEDQLOQKKEISQFNLRLPQOQHETASQTSFPDYNQGTQAVTEEN 469

QY 653 LVLVAQVAAEMRRNGLSLISQYFYQDYKCREMYDKDILMLQI-----GAS 700

DB 470 IASLQKRVVELENEKGALLSSI-ELEELKAENEKLSQITLLAEQNRTGEADREVSEIS 528

QY 701 LMD-PNK-----FLLLVLQRYELAEA 720

DB 529 IVDIANKRSSAEESQDVLNTFSQKHKLSVLLLEMKEAQBEIAFLKLQLOQKRAEEA 588

QY 721 FNKTISTKQDOLIKQY--NTLIEMLQVLIYIGERY----- 755

DB 589 DHEVL---DOKEMKQMEGEGIAPIKMKVLEFDGTGQDFPLMPNEESSLPVAVEKEQASTEHQ 645

QY 756 -----VPGVGNVTKEVMTWREITHLCIEPMPSHA- 785

DB 646 SRTSEISINDAGVELKSTKQDGDKSLSAVPDQCQHQDELE-RLKSQIILELELNFHKAQ 704

QY 786 --IAKNLPENENNETGLENVINKVAFKPGVSGHGVY-----ELKDESL---KDFNMYF- 835

DB 705 EIYKNLDEKAKEISNLNOLIEE---FKNADNNSAFTALSEERDQLLSQVKELSMWTE 761

QY 836 --YHYSKTQHSKAEMQKRRKQENK---DEALPPPPPPFCPAFSKVINLLNCDIMMYI 890

DB 762 LRAQVKQLEMLNLAEAERQRLDYESQTAHDLNLT-----QIHSLSTEAKSKDVKIEV 814

QY 891 LRTVFE-----RAIDTDSNLWTEGMLQMAFH 916

DB 815 LONELDDVQLOFSEOSTLIRLSQLOQNKSEVLEGAERVRHISSKVVELSQALSQKELE 874

QY 917 ILALG--LLEEK--QOLQKAPEE-----EVTFFDYFKASRLG-----SSAMNIOMLLEK 961

DB 875 ITKMDQLLEKKRDVETLQOTIEKDQOVTEISFSFTEKMWQLNEEKFSGLGVIEIKLKEQ 934

QY 962 LKGI-----POLEGQKDMITWLOMPTVKRLREKSLIVATTSSESINKDEITHD 1013

DB 935 LNLLSRAEAKKEQVEEDNEVSSGLKNYDEM-----SPAQGISKEELOHE 980

QY 1014 -----KEAERKRKAEEAARLHRQKIMQMSALOKNFIEPHKLMYDNTSEMPKEDSIMEE 1068

DB 981 FDLKKKENEQRKRLQAAALINKELLQVRSLEE---ELANLKDESKKEIPLSETERGEV 1037

QY 1069 ESTPAVSQVSRITALGPKRGPSVTEKEVLTCLQEE-----EQEVKIENNAWVLS- 1117

DB 1038 BEDKENKEISE-----KCVTSK-----CQEIYVLKOTISEKEVELQHIRKDLEE 1082

QY 1118 --ACVQKSTALTQHRGKPIELSGEALDPL 1144

DB 1083 KLAABEQQALVKQNNQTLQDKTNOIDLL 1111

RESULT 13

A57013

N:Alternate names: endosome-associated protein

C:Species: Homo sapiens (man)

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C:Accession: A57013; S44243

R:Mu, F.T.; Callaghan, J.M.; Steele-Mortimer, O.; Stenmark, H.; Parton, R.G.; Campbell

J. Biol. Chem. 270, 13503-13511, 1995

A:Title: EEAL, an early endosome-associated protein. EEAL is a conserved alpha-helica

A:Reference number: A57013; MUID:95286647

A:Accession: A57013

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: mRNA

A:Residues: 1-1410 <RES>

A:Cross-references: GB:L40157; NID:g1016367; PIDN:AAA79121.1; PID:g1016368

R:Seelig, H.P.

submitted to the EMBL Data Library, April 1994

A:Reference number: S44243

A:Accession: S44243

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-254; C, 256-257, 'LQ', 260-276, 'A', 278-283, 'A', 285-519, 'D', 521-574, 'EQ', 5

A:Cross-references: EMBL:X78998; NID:g475933; PIDN:CAA55632.1; PID:g475934

C:Genetics:

A:Gene: GDB:EEAL

A:Cross-references: GDB:1369996

C:Superfamily: human early endosome antigen 1

C:Keywords: calmodulin binding; endocytosis; metal binding; peripheral membrane prote

Query Match 2.0%; Score 186; DB 1; Length 1410;

Best Local Similarity 18.5%; Pred. No. 0.0033;

Matches 204; Conservative 200; Mismatches 424; Indels 272; Gaps 45;

QY 176 ENSRCPNEEVIVQARKIPPSVIKYVVMETWEEKELPPELQIREKNERYCVLFDNRH 235

DB 275 ELSKGPQEVRYVQELQKLKSSVNLTKNQTLTNLLKKEQDYTKLEKH-----NEES 329

QY 236 HSYDHVIVSYL-ORALDCELAELAHHTTAIDKEGRRAVKAQYAAQCAKEDTKSHSNVS 294

DB 330 VSKNKIQATLHKDKDCCQLQSLR-----SASE-TS 359


```
Db 815 -----LDTERSQAQERDAARQAQAQEGKTALEQQAHEKVENQ-----L 858
QY 892 RTVFERADTDNSLWTEGMLQWAFHILALGLLEKQOQLOKAPEEEEVTFDYHKASRLGSS 951
Db 859 REKWEK-----ERSHQOQELAKALESLEKEMLEMLKEQOQTEMAIOQAQREERETQAE 913
QY 952 AMNIQMLEKLGIPQLEGKDMITWILQMFDTVKRLREKSCLI--VATTSGSESIKDE 1009
Db 914 SALCOM-----QLETEKERSVLETLTQTOQELADASOOLERLQDMKVQKLKEQE 964
QY 1010 IT-----HDKAEKRRKAEARHLRQKIMA---QMSALQKNFIETHKLMYDNTSEMPGKE 1062
Db 965 TTGILQTLQEAQRELK-EAARQHRDDLAALQEESSILQDKWDLQKQVEDLKSQSLVAOD 1023
QY 1063 DS--TME---BESTPAVDYSRIALQPKRGVSTKEVLTCLQCEEQEVQKIENNAMWLS 1117
Db 1024 DSORLVEQVEKLTRETQEYNRQIKELER-----EKASLTLSLMEKEQRLVLQEAD--S 1076
QY 1118 ACVOKSTALTOH---RGKPIELSGE-----ALDPLFMDPLAYGTYTGSCGHV 1162
Db 1077 IQQELSALRODMQBAQGEQKELSQMELLROEVKEADFLAQEAQLLEEL---EASHI 1133
QY 1163 ---MHAYCWQKYFAVQVLSQORIHVDLFDLESGEYLCPLCKSLCNTVPIPILOPQKI 1218
Db 1134 TEQQLRASLWAQEAQAQL-----HLRSTES-----QLEALAAEQ 1171
QY 1219 NSENADALQALLTLARWITQTVLARI 1243
Db 1172 PGNAQAQAQALASLYSALQAALGSV 1196

RESULT 15
T26467
hypothetical protein Y11D7A.14 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 08-Sep-2000
C:Accession: T26467
R:Steward, C.
submitted to the EMBL Data Library, October 1998
A:Reference number: Z20218
A:Accession: T26467
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-1413 <WTL>
A:Cross-references: EMBL:AL032632; PIDN:CAA21588.1; GSPDB:GN00022; CESP:Y11D7A.14
A:Experimental source: clone Y11D7A
C:Genetics:
A:Gene: CESP:Y11D7A.14
A:Map position: 4
A:Introns: 17/3; 62/3; 122/2; 178/3; 200/1; 312/1; 351/3; 387/1; 463/3; 561/1; 583/3; 62
A:Superfamily: myosin heavy chain; myosin motor domain homology
F:35-696/Domain: myosin motor domain homology <NMO>
```

```
Query Match 2.0%; Score 180.5; DB 2; Length 1413;
Best Local Similarity 20.4%; Pred. No. 0.0073;
Matches 187; Conservative 127; Mismatches 350; Indels 253; Gaps 38;

QY 286 IKSHSENVSOHPLHVEVLHSEIMAHQFALRLGSGMKNIM--SYSSDFRQIFCOACLREE 343
Db 336 IKVHDKLIRKNNLAKTLSSASAMAKILYERLFGVIRKCNDAFVSD----- 382
QY 344 PDSENPCLISRLMDAKILKGARKILHELIFSSFFMEMEYKKLFAMEFVKKYQLOQKEY 403
Db 383 -DTESTCLRSRFI---AVLDIAGFEIIEKNSFEQPCINYNEKL--QQFENHMFQAEQ- 435
QY 404 ISDDHRSITISIALSVQMTVPTPLARHLEE-ONVISVITE-----TLLEVLPEY 452
Db 436 -SDYLEEGIKWTVQNFANHLQPTI--DLTEKPMGILSFLEEECVPGNGSEKSLLEKLCN 492
QY 453 LDRNNKF-----NFGYSQDKLGRVYAVICDLKYILISK 486
Db 493 LNSDSFKKSKOTQKQKSTIRHFYVQHYAGEVHYNDGWLDKNRDNVETSVLD-----ILSQ 548
```

Search completed: February 28, 2002, 10:02:40
Job time: 277 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 28, 2002, 09:57:03 ; Search time 19.45 seconds
(without alignments)
2023.561 Million cell updates/sec

Title: US-09-724-126A-2
Perfect score: 9224
Sequence: 1 MADEEAGGTERMEISALPQ.....EYASQETNOMLGFNQLL 1749

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues
Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	156.5	1.7	2710	1	US-08-480-604A-6	Sequence 6, Appli
2	156.5	1.7	2710	2	US-08-405-496A-6	Sequence 6, Appli
3	156.5	1.7	2710	4	US-08-915-136-6	Sequence 6, Appli
4	142.5	1.5	3248	1	US-08-353-700-1	Sequence 1, Appli
5	142.5	1.5	3248	5	PCT-US95-16216-1	Sequence 1, Appli
6	140.5	1.5	1708	1	US-08-493-092-2	Sequence 2, Appli
7	140.5	1.5	1708	1	US-08-508-836A-2	Sequence 2, Appli
8	139.5	1.5	2482	1	US-08-328-254-6	Sequence 6, Appli
9	139.5	1.5	976	4	US-09-104-324B-4	Sequence 4, Appli
10	133.5	1.4	3056	1	US-08-508-836A-8	Sequence 8, Appli
11	133.5	1.4	3056	2	US-08-629-001A-3	Sequence 3, Appli
12	133.5	1.4	3056	2	US-08-874-266-2	Sequence 2, Appli
13	133.5	1.4	3056	4	US-08-642-274D-3	Sequence 3, Appli
14	133.5	1.4	3056	4	US-08-952-127-3	Sequence 3, Appli
15	133.5	1.4	3056	4	US-08-952-014C-3	Sequence 3, Appli
16	132.5	1.4	1276	1	US-08-717-515-8	Sequence 8, Appli
17	131	1.4	1073	4	US-09-541-782-6	Sequence 6, Appli
18	130	1.4	2368	1	US-08-198-446B-15	Sequence 15, Appli
19	130	1.4	2368	2	US-08-870-693-15	Sequence 15, Appli
20	129.5	1.4	712	2	US-08-468-576B-17	Sequence 17, Appli
21	129.5	1.4	712	2	US-08-468-577B-17	Sequence 17, Appli
22	129.5	1.4	712	3	US-08-468-577B-17	Sequence 17, Appli
23	126	1.4	3672	2	US-08-822-445-12	Sequence 12, Appli
24	126	1.4	3801	2	US-08-822-445-10	Sequence 10, Appli
25	125.5	1.4	2154	2	US-08-841-349-4	Sequence 4, Appli
26	124	1.3	2860	2	US-08-826-267-2	Sequence 2, Appli
27	123.5	1.3	3959	2	US-08-970-269A-30	Sequence 30, Appli

28	123.5	1.3	3959	4	US-09-407-562-30	Sequence 30, Appli
29	123	1.3	876	1	US-08-717-515-4	Sequence 4, Appli
30	123	1.3	988	3	US-08-851-843A-69	Sequence 69, Appli
31	123	1.3	988	4	US-08-974-549A-112	Sequence 112, Appli
32	123	1.3	988	4	US-08-854-050-69	Sequence 69, Appli
33	123	1.3	1354	3	US-08-885-871-2	Sequence 2, Appli
34	123	1.3	2101	1	US-08-466-390-4	Sequence 4, Appli
35	123	1.3	2101	1	US-08-470-950-4	Sequence 4, Appli
36	123	1.3	2101	1	US-08-467-781-4	Sequence 4, Appli
37	123	1.3	2101	1	US-08-195-487-4	Sequence 4, Appli
38	123	1.3	2101	2	US-08-483-924-4	Sequence 4, Appli
39	123	1.3	2101	4	US-09-452-294-1	Sequence 1, Appli
40	123	1.3	2101	5	PCT-US93-06160-4	Sequence 4, Appli
41	122.5	1.3	1057	4	US-09-541-782-10	Sequence 10, Appli
42	121	1.3	577	4	US-09-315-794-32	Sequence 32, Appli
43	121	1.3	577	4	US-09-389-341-32	Sequence 32, Appli
44	121	1.3	816	2	US-08-533-306A-6	Sequence 6, Appli
45	121	1.3	816	2	US-08-742-923A-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-08-480-604A-6
; Sequence 6, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,604A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,711
; FILING DATE: 14-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027

```

; REFERENCE/DOCKET NUMBER: OPHD-01763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-480-604A-6

Query Match 1.7%; Score 156.5; DB 1; Length 2710;
Best Local Similarity 18.6%; Pred. No. 0.00041;
Matches 189; Conservative 154; Mismatches 354; Indels 317; Gaps 44;

Qy 283 KEDIKSHENVSQHPHVEV---LHSEIMAHOKFALRLGSMWNKIMSYSSDFRQIFCOAC 339
Db 355 KSEIFSKLENLVSDDLEIKITAFALGVI--NQALISKQGSYLTNLVTEQVKNRYQFLNQH 412

Qy 340 LREEPOSENCLISRLMLDAKLYKGARKILHELIFSSFFMEMEYKKLFAMEFVKYKQL 399
Db 413 LNPAIESDNN-----FTDTTKIFHDSLFNSATAE---NSMFLTKIAPY---L 453

Qy 400 QKEYISDDHRSISITA-----LSVQMTFV-PTL-ARHLIE---EQNVISVITE 443
Db 454 QVGFMPEARS-TISLSPGAYASAYYDFINLQENTIEKTLKASDLIEFKFPENNLSQLTE 512

Qy 444 TLLEVLPEYLDNRKNFQCYSDKLGRYVA-----VICDLKYLISK-PTIWE 492
Db 513 QEINLSWSPQAAKQAFYKIVRYDITGGSUSDNQVDFNKTALDKNYLLNNKIPSNVNE 572

Qy 493 RLRMQFLEGRSLKILTCMQGMEIRRQVGHIEVDPDWEAAIAIQMLKNILLMFQEW 552
Db 573 E-----AGSKNVVHYIIIOGGDDISYEATCNLFKNP--KNSIIIOBNMESAKSY--F 622

Qy 553 CADDELLVAYKECHKAVMRCSTSFSSSKTVVQSCGH-----SLETSYR 599
Db 623 LSDDGESIL---ELNK--YRIPERLKNKEKVKVTFIGHGKDEFNTSEFARLSVDLSNE 676

Qy 600 VSEDVSIHLPLSRTLAGLHVRLSRLGAVSRHLHEFFSDFEFOVELVEYPLRCLVLVAQV 659
Db 677 ISSFLDTIKLIDISPK-----NVEVNLGCG-----NMFSY-DFNVE--ETYPGKLLLSIMDK 724

Qy 660 VA-----EMWRNGLSLISQVFFYQDVKCREMY--- 688
Db 725 ITSTLPDVNKNSTIGANQYEVINSEGRKELLAHSGKWNKKEAIMSDLSKSEYIFDS 784

Qy 689 -----DKDIIIMQIGASLMDPNKFLLLVLQ-----RYELAE 719
Db 785 IDNKLAKSKNIPGLASISEDIKTLLDASVSPDKFPLNNKLNIESSIGDIYIYEKLE 844

Qy 720 AFNKTSTKQDQLIKQYNTLIEMLQVLIIVGERYVPGVGNVTKKEVTWREITHLCIE 779
Db 845 PVKNIHNSIDDLIDFENLL-----ENVSDELYELKKL----- 877

Qy 780 PMPHSAIAKNLPE-----NENNETGLENVINKVATFKKPCVSGHGYVELKDESILKD 830
Db 878 -----NNLDEKYLISFEDISKNNSTYSVRFINK-----SNGESVIVETEKEI-- 919

Qy 831 FNMVYFYHSK-----TQHSKAEHMOKRRKQENKDEALPPPPPEFCFAPAFSKV 878
Db 920 FSKYSEHTKTEISTIKNSIITDVNGNLLDNIQLDHTSQ----- 957

Qy 879 INLLNCIDIMMYILRTVFERAIDTD--SNLWTEGLMQAFHILALGLLEEKQOQLQAPEEE 936
Db 958 VNTLNA---AFFIOSLIDYSSNKDVLNDLSTSVKQVLYAQLFSTGLNTIYDSIQLV--- 1010

Qy 937 VTDFETHKASRLGSSAMN---IQMLLEKLGKIPOLEGOKDMITW---ILQWPDVTVKRLREK 991
Db 1011 -----NLSISNAVNNDPTINVLPITTEGIPTIVSTILDINGLGAATKELLDHDPDLLK 1060
```

```

Qy 992 -----SCLIVATTSGSESISIKNDEITHDKAEKRRKKAEBARLHROKIMQAOMS 1038
Db 1061 ELEAKVGVLAINMSLSIATVASIVGI- GA EVT-----IFLLPIAGISAGIP 1106

Qy 1039 ALOKNFIETHKLMYDNTSEMPCKEDSIMEEESTPAVDYSRIALGPKRGSPVTEKE-VLT 1097
Db 1107 SLVNN-----ELILDHKATVSVVYFNHULSESCKYKGLPKTKEDDKILV 1147

Qy 1098 CILCOEEQEVKIENNAMVLSACVQKSTALTQHRGKPIELSGEALDPLFMDPDILA 1151
Db 1148 PIDDLVISEIDFNNSIKLGTG---NILAMEGSGHTVTGN-IDHFFSSPSIS 1196

RESULT 2
US-08-405-496A-6
; Sequence 6, Application US/08405496A
; Patent No. 5919665
; GENERAL INFORMATION:
; APPLICANT: WILLIAMS, JAMES A.
; TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/405,496A
; FILING DATE: 16-MAR-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01308
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-405-496A-6
```

Query Match 1.7%; Score 156.5; DB 2; Length 2710;
Best Local Similarity 18.6%; Pred. No. 0.00041;
Matches 189; Conservative 154; Mismatches 354; Indels 317; Gaps 44;

Qy 283 KEDIKSHENVSQHPHVEV---LHSEIMAHOKFALRLGSMWNKIMSYSSDFRQIFCOAC 339
Db 355 KSEIFSKLENLVSDDLEIKITAFALGVI--NQALISKQGSYLTNLVTEQVKNRYQFLNQH 412

```

Qy 340 LREPDSENCLISRLMLDAKYLKYGARKILHELIFSSFFMEMEYKKLFAFMEFYKYLQ 399
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 413 LNPAIESDNN-----FTDTTKIFHDSLFNSATAE---NSMFLTKIAPY---L 453
Qy 400 QKEYISDDHRSISITA-----LSVQMTV-PTL-ARHLIE---EONVISVITE 443
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 454 QVGFMPPEARS-TISLSGPGAYASAYDFINLQENTIEKTLKASDLIEFKFPENLSOLTE 512
Qy 444 TLLEVLPEYLDNRNKNFQYSDQKLGRYVA-----VICDLKYILISK-PTIWE 492
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 513 QEINSLSFQASAKYQFVKYVRDYTGSLSDNGVDENKNTALDKNYLLNKNKIPSNVE 572
Qy 493 RLRMQFLGRFSLKILTCMQGMBEIRQVQGHIEVDPDMFAAIAIQMLKNILLMFQEW 552
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 573 E-----AGSKNVVHYIIQLQDDISYEATCNLFKNP--KNSIIQRMNESAKSY--F 622
Qy 553 CACDEBLLVAYKECHKAVRCSTSFSSKTVVQSCGH-----SLETSYR 599
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 623 LSDGESIL-----ELNK--YRIPERLKNKEKVKVTFIGHGKDEFNTSEFARLSVDSLSNE 676
Qy 600 VSDLVSLHLPLSLKLAGLHRLSRLGAVSLRHEFEVFEQVVEVLYEYPLRLCLVLAQV 659
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 677 ISSFLDTIKLIDSPK-----NVEVNLGCG-----NMFSY-DFNVE--ETYPGKLLLSIMDK 724
Qy 660 VA-----EMWRRNGLSLISQVFFYQDVVKCREMY--- 688
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 725 ITSLPDVNKNSITIGANQYEVRIINSEGRKELLAHSKWKINKEAIIIMSDLSSEYIFPDS 784
Qy 689 -----DKDIIMLTQASLMDPNKFLVLVQ-----RYELAE 719
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 785 IDNKLKAKSNIPGLASISEDIIKTLDDASVDPDKFTLNLLKNIESSIGDYIYEKLE 844
Qy 720 AFNKTISTKODDLIKQNTLIEEMQLVLIYVGRYVPGVGNVTKKEVTMRREIHLICIE 779
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 845 PVKNIHNSDDLDIDFNLL-----ENVSDELYELKKL----- 877
Qy 780 PMPHSAIAKLNPE-----NENNETGLENVINKVATEKKPGVSGHGVVYELKDESLKD 830
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 878 -----NNLDEKYLISFEDISKNNSTYSVRFINK-----SNGESVYVETEKEI-- 919
Qy 831 FNMVTFYHSK-----TQHSKAEHMVKRRKQENKDEALPPPPPEFCPAPFSKV 878
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 920 FSKYSEHITREIKSTIKNSIITDVNGNLLDNIQDHTSQ----- 957
Qy 879 INLLNCIDIMYILRTVERAIDTD--SNLWTEGMLQAFHILALGILLEKQOLQKAPEEE 936
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 958 VNTLNA---AFFIOSLIDYSSNKVDLNDLSTSVKVLQYALQFSTGLNTIYDSIOLV--- 1010
Qy 937 VTDFYHKASRLGSSAMN--IQMLLEKLGIPQLEGQKDMTW---ILOMFDTVKRLREK 991
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1011 -----NLSINAVNDYINVLPTTIGIPVSTILDGINLCAATKELLDEHDPPLKK 1060
Qy 992 -----SCLIVATTSSSEIKNDEITHDKEAERKKAERKAAARLHRQKIMAQMS 1038
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1061 ELEAKVGVLAINNLSLSAATVASIVGI-GAEVT-----IFLLPIAGISAGIP 1106
Qy 1039 ALQXNFETHKLMYDNTSEMPGKEDSTMEESTPAVSDYSRIALGPKRGPSVTEKE-VLT 1097
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1107 SLVNN-----ELLHDKATSVVNFYFNHLSSESKYGLPKLTEDDKILV 1147
Qy 1098 CILCOEQEYKIEENAMVLSACVQKSTALTQHRGKPIELSGEALDPLFMDPDLA 1151
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1148 PIDDLVISEIDFNNSIKLGTG-----NILAMEGSGGHTVTGN-IDHFFSPSPIS 1196

```

RESULT 3

```

US-08-915-136-6
; Sequence 6, Application US/08915136
; Patent No. 6290960
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.

```

```

; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESS: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/915,136
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/480,604
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-915-136-6

```

```

Query Match 1.7%; Score 156.5; DB 4; Length 2710;
Best Local Similarity 18.6%; Pred. No. 0.00041;
Matches 189; Conservative 154; Mismatches 354; Indels 317; Gaps 44;

Qy 283 KEDIKSHSNVSOHPLHVEV---LHSEIMAHOKFALRLGSMNKNISYSSDFRQIFCOAC 339
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 355 KSEIFSKLENVSDLEIKIAFALGSKI---NQALISKQSYLNLVIEQVKNRYQFNLQH 412
Qy 340 LREPDSENCLISRLMLDAKYLKYGARKILHELIFSSFFMEMEYKKLFAFMEFYKYLQ 399
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 413 LNPAIESDNN-----FTDTTKIFHDSLFNSATAE---NSMFLTKIAPY---L 453
Qy 400 QKEYISDDHRSISITA-----LSVQMTV-PTL-ARHLIE---EONVISVITE 443
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 454 QVGFMPPEARS-TISLSGPGAYASAYDFINLQENTIEKTLKASDLIEFKFPENLSOLTE 512
Qy 444 TLLEVLPEYLDNRNKNFQYSDQKLGRYVA-----VICDLKYILISK-PTIWE 492

```

Db 513 QINLSWSDQASAKYQFKYVRYDTGGSLSDNGVDNFKNKTALDKNYLLNNKIPSNVVE 572
QY 493 RLRMFQLEGRSFKILTCMQMEEIRROVQGHIEVDPOWEAATAIOMOLKNILLMFQEW 552
Db 573 E-----AGSKNVVHYIIOGGDISYEATCNLFKSNP--KNSIIQRNNESAKSY--F 622
QY 553 CACDELLLVAYKECHKAVNRSTSFSSKTVVQSCGH-----SLETKSYR 599
Db 623 LSDGESIL-----ELNK--YRIPERLKNKEKVKTIGHGKDEFNTSEFARLSVDSLSNE 676
QY 600 VSEDVSLHPLSRTLGLAGHLVRLSRLGAYSLHEFVSFEDFQVEVLVEYPLRCLVLVAQV 659
Db 677 ISSFLTDIKLIDSPK-----NVEVNLGCG-----NMFYSY-DFNVE--ETYPGKLLLSIMDK 724
QY 660 VA-----EMVRRNGSLISGVFYQVQVCKREMY--- 688
Db 725 ITSTLPDYNKNSITIGANQYEVINSEGRKELLASHGKWKINKEEATMSDLSKEYIFPDS 764
QY 689 -----DKDIIMLOIGASLMDPNKFLLLVLO-----RYELAE 719
Db 785 IDNKLKAKSNIPGLASISDIKTLLDASVSPDTFKILNKLNIETSSIGDYIYYEKL 844
QY 720 AFNKTISTKQDLIKOYNLTIEBMLQVLIYIGERVVPGVGNVTKREVTMRIIHLCLIE 779
Db 845 PVKTIHNSIDLDIFNLL-----ENVSDELYELKLL----- 877
QY 780 PMPHSAIAKNLPE-----NENNETGLENVINKVATFKPGVSGHGVYELKDSLSKD 830
Db 878 -----NNLDEKYLISFEDISKNNSTSVRFINK-----SNGESVYVETEKEI-- 919
QY 831 FMYIYTHYSK-----TOHSAKABHMQKRRKQENKDEALPPPPPPFCPAFSKV 878
Db 920 FSKYSEHITKEITKNSIITDVGNNLLDNIQLDHTSQ----- 957
QY 879 INLLCNDIMYILRVTFERAIDTD--SNLWTEGMLQMAHIALGLLEKQLOKAPBEE 936
Db 958 VNTLNA-----AFFIQSLIDYSKNDVLDNSTSVKQVLAQLFSTGLNTIYDSIQLV--- 1010
QY 937 VTFDFVHRASRLGSSAMN--IOMLEKLGIPOLEGQKDMITW---ILOMFDTVVKRLREK 991
Db 1011 -----NLISNAVNTINVLPTITGIPVISTILGINLGAIAKELLDEHDP LKK 1060
QY 992 -----SCLIVATTSGESIKNDEITHDKEAERKKAABAARLHRQKIMAOMS 1038
Db 1061 ELEAKGVGLAINMSLSIAATVASIVGI-GAEVT-----IFLLPIAGISAGIP 1106
QY 1039 ALOKFNIEHKLMDYNTSEMPGKEDSIMSEESTPAVSDYSRIALGPKRGPSTVEKE-VLT 1097
Db 1107 SLVNN-----ELILHDKATSVVNYFNHLSSESKYGLPKTEDDKILV 1147
QY 1098 QTLCOEEOBKVIENAMVLSACVOKSTALTQHRGKPIELSGEALDPLFMDPDLA 1151
Db 1148 PIDDLVISEIDNNNSIKLGTG-----NILAMEGGSGHTVTGN-IDHFFSSPSIS 1196

RESULT 4

US-08-353-700-1
; Sequence 1, Application US/08353700
; Patent No. 5599919
; GENERAL INFORMATION:
; APPLICANT: YEN, TIMOTHY J.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A
; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESS: DANN, DOREMAN, HERRELL AND SKILLMAN
; STREET: 1601 MARKET STREET, SUITE 720
; CITY: PHILADELPHIA
; STATE: PA
; COUNTRY: USA

; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/353.700
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: REED, JANET E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: HUMAN
; US-08-353-700-1

Query Match 1.5%; Score 142.5; DB 1; Length 3248;
Best Local Similarity 18.2%; Pred. No. 0.01;
Matches 302; Conservative 245; Mismatches 601; Indels 515; Gaps 76;

QY 166 HEPGRAGTIFKNSRCPLNEE---VIVQARKIFPSVIKVVVEMTWEE---KELPPELQ 218
Db 304 HEKEMGOVNFQELQLEKAKVELIEKEVKNKCRDELVRTTAQYDQASTKYTALEOK 363
QY 219 IREKNEYCVLFNDE--HHSYDVHYISLQALDCELAQAHLHTTAIDKEGRRVAKAGAY 276
Db 364 LKLTEDLSQORNAESARCSLEQKIKEKEFEQELSROORSFOTLQECIQ--MKARLT 422
QY 277 AACQEAEDTKSHSENSVQHPHLEVLHSEIMAHOKFALRLGSGWNKIMSYSSDRFQIFC 336
Db 423 QELQAKRN-----MH-NVLOAELD-----KLTSVQKQLENNLEEFKQKLC 461
QY 337 QACLREEPDSENPCLSRLMLDALKYKGARKILHELIFSPFEMEKYKLFAMEFVKY 396
Db 462 RA-----EQAFOASQIKENELUR--SMEEMKKE 487
QY 397 KQLQKEYISDDHDSISITALSVMFTVP---TLARHLIEEQNVISVITETLLEVLPEYL 453
Db 488 NNLK---SHSEQAKAREVCHLEAEKLNKQCLNOSONFAEEMKAKNTSOETMLRDLQE-- 542
QY 454 DRNKNFQGYSDQKLGKRVYAVICDLKYLISKPTIWTLELRMQFLEGFRSLKILTQMQ 513
Db 543 ---KINQOENSL-TLEKLKLAVADLE-----KQDCSQ 571
QY 514 GMEIRQVGOHIE-----VDPDWEAAI-AIQM-----OLKNILLMFQEWACDEE 558
Db 572 DLLKRE---HHIEQLNDKLSKTEKESKALLSALELKKKEYEELKEKTLFSCWSENEK 628
QY 559 LLLVAYKECHKAVNRSTSFSSKTVVQSCGHSLETSKSYRVSDELVSITHLP---LSRTL 615
Db 629 LL-----TQMESEKENLQSKINHLETCLKTQIKSHYENRVRTLEMDRENLSVEI 679
QY 616 AGHL-----VRLSRLGAVSRLHEFVSFEDFQVEVLVE-YPLRCLVLVAQVVAEMMR 665
Db 680 RNLHNVLDSSKSVETQKL-AYMELOKQAEFSQDKHQKEIENNKLUKTSQLTQVED---L 735
QY 666 RNLGLSLISQVYFYQDVKCREMYD-----KDIIMLOIGASLM---DPNKFLLLVQRYEL 717
Db 736 EHKIQLLSNETMDKD-RCYQDLHAIEYSLRDLKSK-DASLVTNEDHORSLLAFDQOPAM 793


```

QY 718 AEFNKTI-----STKDODLI-----KOYNTLIEML 744
DQ 794 HHSFANIIEGOSMPSECRLEADQSPKNSAILQNRVDSLEFSLESQKQMSDLQOC 853
QY 745 QVLIYIGERYVPGVNTVKEVTMRILHLLCI-----EPMHP-SATAKNLP 791
DQ 854 EELVOIKGE-----IENLAKAEQMHOSFVAETSORISKLOEDTSAHONVVAETLS 904
QY 792 ENNETGLENVINKVAT-----FKPGVSGHGVYELKDESLKDFNMFYHYSKTQHS 844
DQ 905 ALNKEKELQLLNDKVETEQAIEQLKK---SNH-----LLEDSLKELQLL-----S 948
QY 845 KAEHMQKRRKQENKDEALPPPPPEFCFAPFSKVINLLNCDIMMYILRTVFERAID--TD 902
DQ 949 EPLSLEKKE-----NSSIISL-----NKREIEELTQ 974
QY 903 SNLWTEGMLQAFILALGALLEEKQOLQKAPPEEVTDFYHKASRLGSSAMNIQMLEKL 962
DQ 975 EN---GTLKE-----INASLQERNMLIQKSE---SFANYIDEREKSTISELSQYKQKEL 1023
QY 963 KGIPOLEGQKDWITWLOMFTVRLREK-SCLIVATTSGSESINKD---EITHDKKA 1017
DQ 1024 ILLQCEETGNAYEDLSOKYKAAQKNSKLECLNNECTSLCENRKNLEQLKEAFKQHQ 1083
QY 1018 ERKRAEAARLHROKIMQMSALQKNFTETHKLMYDNTSEMPGKEDSIMEESTPAVSQY 1077
DQ 1084 EPLTKLAFABERNQNLMELETVOO---ALRSEMTDNQNNKSEAGLQKQELMT----- 1134
QY 1078 SRIALGPRGSPVTEKEVLTICLOEEOEVKIENNAMVLSACVQKSTALTQHRGKPIELS 1137
DQ 1135 -----LKEQNKQKQEV-NDLQENELQKVMK-----TKHE----- 1165
QY 1138 GEALDPLFMDPLAYGTYTGSCHGMHVAQKQYFEAVOLSSQRIHVDLPDLESEYLC 1197
DQ 1166 -----CONLESEPIRNSVKER-----ESE----- 1184
QY 1198 PLKSLCNTVPIPILOPKINSENADA-LAQLLTLAR-----WIQT 1238
DQ 1185 ---RQCN-FKQMDLEVKETSLDSYNAQLVQLEAMLNKELKLOESKEKELQHELOT 1240
QY 1239 VIARISGYNIRHAKENPIPIFFNOGMDSTLEPHSILSFGVSESSIKEMVILFA 1298
DQ 1241 IRGDLTSLNLDMSQSE-----ISGLKDCIDEABEKYISGPHSLSTSONDAHLCQSLQ 1294
QY 1299 TTIYRIGLKVPPDERPRVPMLTWSTCAFTQATENLLGDECKPLFGALQNRQHNKLAL 1358
DQ 1295 TTMNKLN-----ELEKICEILOAEKYELVLTNLSRSECITAT 1332
QY 1359 MOFAVAORITCPQVLIQKHLVRLSVLPNIKSEDTPCLLSIDLPHVLVGAVALFAPSLVW 1418
DQ 1333 RKMA-----EVEGKLLNEV--KLNDDSG--LLHGELVEDIPGEGF--EQPNE 1375
QY 1419 DDPVLDPSVSSVSNHLYLPHLITMAHMLQILITVDTGLPLAQVQDESEAH-SASSFF 1477
DQ 1376 QHPVSLAPLDESNSYEHLLTSDKEVQMHFAEL---QEFSLSLQSBKILHDOHCQMSKM 1432
QY 1478 AEISQYTSG-----SIGCDIPGWY-----LWVSLKNGITPYLRCAAL----- 1514
DQ 1433 SELQTYVDSLKAENLVLTNIRNFQGDVLVKEMQLGLEGLVPSLSSSCVPSSSLSLGD 1492
QY 1515 ---FFHYLLGVTPPEELHTNSAEGVSA-----LCSYLS-----LPTNLF 1552
DQ 1493 SSFYRALLEQTGDSMLLSN-LEGAVSANQCSQVDEVFCSLSQTYVDSLKAENLVLTNIR 1550
QY 1553 LFQEWYDTR-----PLLRQWCA-DPALLNCLKQKNTVVVYPRKRNLSLIELPDD 1600
DQ 1551 NFQ--GDLVKEMQLGLEGLVPSLSSSCVPSSSLSLGDSSFY-----RALLEQTGD 1601
QY 1601 YSCLLNQASHRCRPRSDDERKHPVLFCGAILCSQNICQOE 1643
DQ 1602 MSLLSNLEGVVANQCSQVDE-----VFCSS-LQEENLTRKE 1636

```

RESULT 5

```

PCT-US95-16216-1
; Sequence 1. Application PC/TUS9516216
; GENERAL INFORMATION:
; APPLICANT: Yen, Timothy J.
; APPLICANT: Rattnner, Jerome B.
; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
; TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/16216
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA: US 08/353,700
; FILING DATE: 09-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
PCT-US95-16216-1

```

```

Query Match 1.5%; Score 142.5; DB 5; Length 3248;
Best Local Similarity 18.2%; Pred. No. 0.01;
Matches 302; Conservative 245; Mismatches 601; Indels 515; Gaps 76;

QY 166 HEPRAGTTKNSRCPLNEB-----VIVQARKIPSPVVIKYYVEMTWEE---KELPPELQ 218
DQ 304 HEKEMGQVKNKFOFLOLQLEKAKVELIEKVKLNKCRDELVRTTAQYDQASTKYTALEQK 363
QY 219 IREKNERYCVLPNDE--HHSYDHVYISLQALDCELAELQHTTATDKEGRRVAKAGAY 276
DQ 364 LKLTEDLSQORQNAESARCSLEQKIKEKEFEQELSRQORSQTLDDQECIO--MKARLT 422
QY 277 AACQEAKEKITKSHNSVNSQHPHVEVLHSEIMAHQFALRLGSMNKMISYSSDFRQIFC 336
DQ 423 QELQQAQN-----MH-NVLAELD-----KLTSVKQOLENNLEEFKQKLC 461
QY 337 QACLREEDPSENCLISRLMLWDAKLYKGARKILHELIFSFFMEMEYKKLFAFMEFYVY 396
DQ 462 RA-----EQAFQASQIKENELRR--SMEEMKKE 487
QY 397 KQLOKEYISDDHRSISITALSVQMTVP---TLARHLIEEQNVISVITETLLEVLPEYL 453
DQ 488 NNLLK---SHSEOKAREVCHLEAELKNIKQCLNQSQNAFEMAKAKNTSQETMLRDQOE-- 542
QY 454 DRNNKFNFGYSDKLGRRVAVVICDLKYILLISKPTIWTERTLRMQFLGFRSFLKILTCMQ 513

```

Db 543 -----KINQENSL-TLEKLVADLE-----KQDCSQ 571
QY 514 GMEIRVQGHTE-----VDPWEAAI-AIQM-----OLKNILLMFOEWACADEE 558
Db 572 DLLKRE---HHTEQLNDKLSKTESKALLSALELKKEVEELKEKTLFSCWKSNEK 628
QY 559 LLLVAYKECHKAVMRCSTSFISSTKVQSCGHSLETKSVRSVSDLSIHLPL---LSRTL 615
Db 629 LL-----TOMSEKENLOKINHLETCLTKQIKSHENVRVLTMDRENLSVEI 679
QY 616 AGLH-----VRLSRGAVSRLEHFEVSFDFQVBLVE-YPLRCLVLVAOVAEMWR 665
Db 680 RNLHNVDSKSEVEYQKL-AYNELOQKAFFSDQKHOKETIENMCLTKTSQLTQGVED---L 735
QY 666 RNLGLISQVYFYQDYKCREMYD-----KDIIMLQIGASLM---DPNKELLVLVQRYEL 717
Db 736 EHKQLLSNEIMDKD-RCYQDLHAESLRLDKSK-DASLVTNEDHORSILAFDQOPAM 793
QY 718 AEFNKTI-----STKDQDLI-----KQYNTLIEML 744
Db 794 HHSFANIIGEGSMPSESRSECRLEADQSPKNSAILQNRVDSLEFSQKOMSDLQKQC 853
QY 745 QVLIYIVGRYRVPVGVNVTKEEYTMREI IHLICI-----EPMPH-SAIKPNP 791
Db 854 EELVQKGE-----TEENLMKAEQMHQSFVAETSQRIKLOEDTSAHQNVVAETLS 904
QY 792 ENENNETGLENVINKVAT-----FKPGVSGHGVYELKDESILKDFNMYFYHYSKTQHS 844
Db 905 ALENKEKELQLNDKVETEQAETQLKK---SNH-----LLEDSILKEQLL-----S 948
QY 845 KAEHMOKKRRKQENKDEALPPPPPECPAFKVINLLNCIDIMMYILRTVFERAID--TD 902
Db 949 ETLSLEKKE-----MSSIISL-----NKREIEBELTQ 974
QY 903 SNLWTEGMLQMAFHIALGILLEKQOOLAKAPEEVTDEYHKASRLGSSAMNQLLEKL 962
Db 975 EN-----GTUKE-----INASUNQKMLQKSE---SPANTIDREKSISELSQYQKEL 1023
QY 963 KGIPOLEGOKDMITWILQFDTVKRLREK-SCLIVATTSGSESINKD-----EITHOKEKA 1017
Db 1024 ILLQRCETGNAYEDLSQYKAAQERNKLECLNNECTSLCENRKNLEQLEKFAKEHQ 1083
QY 1018 ERKRAEALRHQKIMAOXSALQKFIETHKLMYDNTSEMPCKEDSIMEESTPAVS DY 1077
Db 1084 EFLTCLAFAEERNQNLMELETVQO---ALRSEMTDNQNNKSEAGGLKQEI MT-----1134
QY 1078 SRTALGPKRGPSVTEKEVLTCLICEQEOEVKIENNAVLVSACVQKSTALTOHRGKPIELS 1137
Db 1135 -----LKEEQNKQKEV-NDLLOENQELMKVMK-----TKHE-----1165
QY 1138 GEALDPLFMDPDLAYGTYTGTSCGHVMHVCWKYFEAVQLSSQORIHVDLFDLESGEYLC 1197
Db 1166 -----CONLESEPIRNSVKER-----ESE-----1184
QY 1198 PLCKSLCNTVPIPILOPQKINSINADA-LAQLLTLLAR-----WIQT 1238
Db 1185 ---RNQC-FKQPMQLEVEKISLDSYNAQLVQLEAMLRNKELKLOEKEKECLOHELQ 1240
QY 1239 VLARISGYNIRHAKGNPIPIFNQMGDSTLEFHSILSPGVESSIKYSIKEMWILFA 1298
Db 1241 IRGDLETSNLQDMQSEF-----ISGLKCEIDAEEKYISGPHELSTSQDNNAHLQCSLQ 1294
QY 1299 TTYIRGLKVPDPDRPRVPMLTWSTCAFTIOAIENLLGDEGKPLFGALONRQHNGLKAL 1358
Db 1295 TTNKLN-----ELEKICEILOKEKVELVTELNDNRSECIAT 1332
QY 1359 MQFAVAQRITCPQVLQKHLVRLLSVLPNPKSEDPTCLLSIDLPHVLVGAVALFAPS LYW 1418
Db 1333 RKWA-----EEVGKLLNEV-KILNDDSG-LLHGLVEDIPGGEFG-EQPN 1375
QY 1419 DDPVDLOPSVSSSYNHLFIETHWAHMLQIILLTVDTPGLPAQVQEDSEAH-SASSFF 1477
Db 1376 QHPVSLAPLDESNSYEHLTLSDRKQVQMHFAEL---QEKFLSLQSENKILHDOHCQMSKM 1432

QY 1478 AEISOYTSG-----SIGCDIPGWY-----LWVSLKNQITPYLRCAAL-----1514
Db 1433 SELOTYVDSLKAENLVLTNLRNFQDVLVKEMQLGLEGLVPSLSSCVPDSSSLSLGD 1492
QY 1515 --FHYLLGVTPPELHTNSAEGEYSA-----LCSVLS-----LPTNLEL 1552
Db 1493 SSFYRALLEQDGMSSLN-LEGAVSANOCSDVEFCFSLQTYVDSLKAENLVLTNLR-R 1550
QY 1553 LFQEVWDTVR-----PLQWCA-DPALLNCLKOKNTVVRYPRKRNLSLIELPDD 1600
Db 1551 NFQ---GDLVKEMQLGLEGLVPSLSSCVPDSSSLSSGDSFY-----RALLEQTD 1601
QY 1601 YSCLLNQASHFCRPSADDERKHPVLCIFCGAILCSQNICCOE 1643
Db 1602 MSLLSNLEGVVSVANOCSDVE-----VFCSS-LOEENLTRKE 1636

RESULT 6

US-08-493-092-2
; Sequence 2, Application US/08493092
; Patent No. 5728807
; GENERAL INFORMATION:
; APPLICANT: Shiloh, Yosef
; APPLICANT: Taglie, Danilo A.
; APPLICANT: Collins, Francis S.
; TITLE OF INVENTION: Ataxia-Telangiectasia Gene
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Reising, Ethington, Barnard & Perry
; STREET: P.O. Box 4390
; CITY: Troy
; STATE: Michigan
; COUNTRY: US
; ZIP: 48099
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/493,092
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kohn, Kenneth I.
; REGISTRATION NUMBER: 30,955
; REFERENCE/DOCKET NUMBER: P-310 (TAV)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (810) 689-3500
; TELEFAX: (810) 689-4071
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1708 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-493-092-2

Query Match 1.5%; Score 140.5; DB 1; Length 1708;
Best Local Similarity 17.9%; Pred. No. 0.0053;
Matches 260; Conservative 197; Mismatches 469; Indels 529; Gaps 68;
QY 319 SWMKINYSDDFRQIFCOACLR-----EPDSENP-----CL---ISRLMLWD 359
Db 509 SWRNLSTHV---QGFETCLRHFQSRTSTTPANLDBSEHFFRCCLDKKKSQRTMLAV 564
QY 360 AKLYKGARKILHELIFS-SFEMEMEYKKL-----FAMEFVKYKQLOKEVISDD 407
Db 565 VDYMRROKRRSSGTIFNDAPFWDLNLNLEVAKVAOSCAHFTALLYAEIY--ADKXMDQ 622
QY 408 HDRSISITALSVQMTPTVLARHLIEQNVISVITETLLEVLPEYLDNRNKNFQGSOD 467

Db 623 EKSLAFEGS-QSTTSSUSEKSEETGI--SLQDLLLLLEIYRSIGEDSLYCGG----- 675
Qy 468 KLGVRVAVICLKVILISKPTIWERLMOFLEGRSFLKILTCMOGMEETROVGO-HI 526
Db 676 --GKWLQPIELR--TYEHEAMWKAALVYDLE---TAIPSTQAGIIQALQNLGLCHI 728
Qy 527 EVDPDWEAATAIQMLKNILMFOEWCACDELLVA-----YKECHKAYMRC-STSF 579
Db 729 -----LSVLKGLDYENKDWCPLEELHYQAARWQMDHCTSVSKEVEGTSYH 777
Qy 580 SSKTVVQSCG-----HSLFKTSRVSDLSIHLPLSRTLAGLHVRLSRLGAVR 631
Db 778 ESLYNALQSLDRDEFSTFYESLKYARVKEVECMKRSL--ESVYSLYPTLSRLQAI 834
Qy 632 HEFVSFEDFOV-----EVLVEYPLRCLVLVAQVVAEMWRNGLSLISQVFFYODV--- 681
Db 835 ESIGELFSRVTHQLSEVYIK-----WQHSOLLKDSDFSQEPIMAL 878
Qy 682 -----KCREMYDKDIIMLQICASLMDPNKFLLLVLQRYELAEAFNK 723
Db 879 RTVILLELMEKEMDSQRECIKDILTKHLVELSI-----LARTFKN 919
Qy 724 T-ISTKODLIIKOYNTL-----IEMLQVLIYIGERYVPGVGNVTKKEVTMRII 774
Db 920 TQLPERAIFQIKQNSVSCGVSEWQLEE-AQVF-----WAKREQSLASLIL 964
Qy 775 LLCTEPHPSAIKN-----LPENENNETGLENVINKVATPKPGVSGHGVY- 821
Db 965 KOMIKKIDASCAANNPSLKLYTCLRCVGNWLAETCLENPAVIMQTYLXAVEVAGNYD 1024
Qy 822 -----ELKDESLKDFNMYFYHYSTQHSKAEMHOK-----KRRKOE----- 857
Db 1025 GESSDELNRGKMAF--LSLARFSQTQRIENYMKSEFENKQALLKRAKEEVGLLRHK 1083
Qy 858 -----NKDE-----ALPPPPPPFCPAFSKVINLLNCDIM-----MYILRTV 894
Db 1084 IQTNRYTVKQVORELEDELALRALKEDKRPLCKA---VENYINCLLSGEEHDMWVRL- 1139
Qy 895 FERAIDTSLNLTETGLMOMAPHIALGLLEKQOLKAPESVETFD-----YHKASRLGS 950
Db 1140 -----CSLWLENS-----GVSEVNGMKRDMGKIPITYKFLPLMYQLAARMGT 1181
Qy 951 SAMNQMELLEKLIGPOLEGOKMITWLOMTQVRLKREKSLIVATTSSESSEKINDEI 1010
Db 1182 KMWG-----GLGPEVLNLLISRI-----SMDHPHHTLFIILANAN-----RDEF 1223
Qy 1011 THDKERAKR-----KAEAA-----RLHRQKIMQMSALQKNFI--- 1045
Db 1224 LTKPEVARRSRITKNVPKQSSQLDEDRTEAANRIICTIRSRPQWRSVEALCDAYILA 1283
Qy 1046 -----ETHKLMYDNTSEMPCKEDSIWEESTPAV-----SDYS-----R 1079
Db 1284 NLDATOMTKQRGINIPADQPIITLKNLEDVVPVPTMEIKVDHTGEYNLVTIQSKFAFR 1343
Qy 1080 IALGPKRGPSVTEKEVLTCI-----LCOEQOEVEKLEN 1111
Db 1344 LAGG-----VNLPKIIDCVSGSGKERRQLVGRDDLRQDAVMQVQFQMCMTLLQRNFT 1397
Qy 1112 NAMVLSACVOKSTALTQHRG-----KPIELSGEALDPLFMDPDLAYGTYTGSCGHVM 1163
Db 1398 KRKLITICTYKVPVLSQSGVLEWCTGTVPFI---GEF---LVNEDGAHKRYRPNDFSAP 1451
Qy 1164 HAVCWQYFAVQALUSSOORIHVDLFDLESGEYLCPLCKSLCNTVPIPILOPQKINSENA 1223
Db 1452 Q--CQKKMVEYQKKSFEKEVEFM-----DVCQNEQPVF-----RY 1485
Qy 1224 DALAQILTLARWIQTVLA-----RISGVNIRHAKGENPIPIFFNOGMDSTLEPHSI 1275
Db 1486 FCMKFKLDPAIWFKRRLAYTRSVATSSIVGYL-----GLGDRHVQ----- 1526
Qy 1276 LSFGEVSSIKYSNSIKEMVILFATTIYRIGLKVPPDERDP-----RVPMILT 1321

Db 1527 -----NLLINEQSAELVHIDLGVAFEQCKILPTBETVPFRLTRDIVDGMGITVEGVF 1579
Qy 1322 WSTCAFTTQAIEN-----LLGDEGKPLFGALQNRHNGLKALMQFAVAQRITCP 1370
Db 1580 RCCCKETHVMNSQETLLTIVEVLLYD---PLF-----DWTMNPALKALY---LQOR---P 1626
Qy 1371 QVLIQKHVLRLLSVVLPNIKSEDTPCLLSIDLFHVLVGAFLAPFSLYWDPPVDLPQSSVS 1430
Db 1627 EDETELH-----PTLNADDQECKRNL-----SDID 1651
Qy 1431 SSYNHLYFLHLLTMAHMLQIILLTVDLGLPLA---QVQ---EDSEEAHSASSFAEISQYT 1484
Db 1652 QSFQDVAERVLMLRQLEK---GVEEGTVLSVGGQVNLIIQQAIDPKNLSRLF----- 1701
Qy 1485 SSGICGDIPIGWYLV 1499
Db 1702 -----PGWKAWV 1708

RESULT 7
US-08-508-836A-2
; Sequence 2, Application US/08508836A
; GENERAL INFORMATION:
; APPLICANT: Shiloh, Yosef
; APPLICANT: Tagle, Danilo A.
; APPLICANT: Collins, Francis S.
; TITLE OF INVENTION: Ataxia-Telangiectasia Gene
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Reising, Ethington, Barnard & Perry
; STREET: P.O. Box 4390
; CITY: Troy
; STATE: Michigan
; COUNTRY: US
; ZIP: 48099
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/508,836A
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Kohn, Kenneth I.
; REGISTRATION NUMBER: 30,955
; REFERENCE/DOCKET NUMBER: P-313 (TAU)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (810) 689-3500
; TELEFAX: (810) 689-4071
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1708 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-508-836A-2

Query Match 1.5%; Score 140.5; DB 1; Length 1708;
Best Local Similarity 17.9%; Pred. No. 0.0053;
Matches 260; Conservative 197; Mismatches 469; Indels 529; Gaps 68;

Qy 319 SWNNKIMYSDFRQIFCOACLR-----EPDSENP-----CL---ISRMLWD 359
Db 509 SWRNLSTHV---QGFTTSLRHFQTSRSTTPANLDSESEHFFRCCLDKKQORTMLAV 564
Qy 360 AKLYKGARKILHELIFS-SFFMEMEYKKL-----FAMEFYKYLQKQKEYISDD 407
Db 565 VDMRRQRKPSSTGTFINDAFWLDNLNLEVAQVACAAHFTALLYAEIY--ADKKSMDQ 622

Db 2488 -----CSWLLENS-----GVSEVNGMMKRGDMKIPYKFLPLMYQLAARMGT 2529
Qy 951 SAMTQIMLEKLGIPQLEGOKMITWILQMFDTVKRLREKSLIVATTSSESINKDEI 1010
Db 2530 KMG-----GLGFHEVLNLLISRI-----SMDHPHHTLFILALANAN-----RDEF 2571
Qy 1011 THDKEKAERKR-----KAEAA-----RLHRQKIMQMSALQKNFI--- 1045
Db 2572 LTKPEVARRSRITKNVPKQSSOLDREDTEAANRIICTIRSRPQMVRSVEALCDAIILA 2631
Qy 1046 -----ETHKLMYDNTSEMPKGEDSIMEESTPAV-----SDYS-----R 1079
Db 2632 NLDATQMTORGINIPADQITLKNLEDDVVVPTMEIKVDHTGEYGNLVITQSFKAERF 2691
Qy 1080 IALGPKRGPSVTEKEVLTCI-----LCQEEQEVKTEN 1111
Db 2692 LAGG-----VNLPIIDCVSGDKERRQLVKGRDDLRODAVMQOVQMCNTLILQRTET 2745
Qy 1112 NAMVLSACVOKSTALTOHRG-----KPIELSGEALDPLFMDPDLAYGTYTSCGHVM 1163
Db 2746 RKRLTICTYKVPPLSQRSVLEWCTGTVP-----GEF-----LVNNEDEGAHKRYRPNDFSFA 2799
Qy 1164 HAVCWQKYFEAVQLSSOORIHVDLFDESGLPLCKSLCNTVPIPILOPQKINSENA 1223
Db 2800 Q--COKMMEYQKKSFEEKVEVFM-----DVCNQFPVF-----RY 2833
Qy 1224 DALAQLTLARWITVLA-----RISGYNIRHAKGENPIPIFFNOGMDSTLEFHSI 1275
Db 2834 FCMKEFLDPAWFKRLAYRSVATSSIVGIL-----GLGDRHVQ----- 2874
Qy 1276 LSGVESIKYSIKEMVILFATYIRIGLVKPPDERDP-----RVPMILT 1321
Db 2875 -----NILNEQSAELVHIDLGVAFEQGLIPTPETVPERLTDIVDGMGTVGEVGF 2927
Qy 1322 WSTCAFTIOAIEN-----LLGDEGKPLFGALQNRHNGLKALMQFAQRITCP 1370
Db 2928 RRCEKTEMEVRNRSOETLLATIVELLYD---PLF---DWTMPLKALY---LQOR---P 2974
Qy 1371 QVLIQKHLVRLSVVPLNPKSEDPCLLSIDLFLVLCVAVLAPFLSLYWDVDPVLOPSSVS 1430
Db 2975 EDETELH-----PTLNADDOECRNL-----SDID 2999
Qy 1431 SSYNHLXFLHITMAHMLQILLTVDTGPLA---QVQ---EDSEEAHSASSFFAEISQYT 1484
Db 3000 QSEPKVAERVLMLRLOELK---GVEEGTVLSVGQVNLIIQOAIIDPKNLSRLF----- 3049
Qy 1485 SGSGICDIPGWLWV 1499
Db 3050 -----PGWKAWV 3056

RESULT 11

US-08-629-001A-3
; Sequence 3, Application US/08629001A
; Patent No. 5858661
; GENERAL INFORMATION:
; APPLICANT: Shiloh, Yosef
; TITLE OF INVENTION: ATAXIA-TELANGIECTASIA GENE AND ITS
; TITLE OF INVENTION: GENOMIC ORGANIZATION
; NUMBER OF SEQUENCES: 139
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kohn & Associates
; STREET: 30500 No. 5858661thwestern Hwy.
; CITY: Farmington Hills
; STATE: Michigan
; COUNTRY: US
; ZIP: 48334
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/629,001A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kohn, Kenneth I.
; REGISTRATION NUMBER: 30,955
; REFERENCE/DOCKET NUMBER: 2290,00032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (810) 539-5050
; TELEFAX: (810) 539-5055
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3056 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; US-08-629-001A-3

Query Match 1.4%; Score 133.5; DB 2; Length 3056;
Best Local Similarity 17.9%; Pred. No. 0.061;
Matches 260; Conservative 196; Mismatches 470; Indels 529; Gaps 68;

Qy 319 SWNMKMSYSSDFRQIFCOACILRE-----EPDSENP-----CL--ISRMLWD 359
Db 1857 SWRNLLSTHV---QGFTTCLRHFSQTSRSTTPANLDSSEHFFRCCLDKKRSORTMLAV 1912
Qy 360 AKLYGKARGKILHELIFS-SFFMEMEYKKL-----FAMEFVKYKQLOKEYSDD 407
Db 1913 VDMRRQKRPSSGTTFNDAFWLDLNLVLEKVAQSCAAHTALLYAEIY--ADKKSMDQ 1970
Qy 408 HDRSITALSVMQFTVPTLARHLIBEQNVISVITETLLEVLPEYLDNRNKNFQGYSD 467
Db 1971 EKSLAFEEGS-QSTTSSLSSEKSEETGI--SLQDILLLEIYRSIGEPDPSLYCGG--- 2023
Qy 468 KLGRVYAVICDLKYLISKPTTWTERLWQFLEGFSFLKILTCMQGMEIRQVQ--HI 526
Db 2024 --GKMLQPITRL--TYEHEAMWGKALVTYDLE--TAIPSSSTRQAGITQALQNLGCHI 2076
Qy 527 EVDPDWEAAIAIQOMKILLMFOEWCACDELLVA-----YKECHKAVMRC-STSEI 579
Db 2077 -----LSVYLGKLDYENKMCPELEELHYQAANRNMOWDHCTSVSKEVEGTSYH 2125
Qy 580 SSKTWTQSCG-----HSLETYSYVSEDLVSIHLPLSRTLGLHLVRLSRLGAVSRL 631
Db 2126 ESLYNALQSLRDREFSTFYESLKYARVKEVEEMCKRSL---ESVSYLPTLSRLQALIGEL 2182
Qy 632 HEFVSFEDFQV-----EVLVEYPLCLVLVAQVAVEMRRNGLSLISQVFFYQDV----- 681
Db 2183 ESIGELFSRSVTHROLSEYIK-----WQKHSQLLKDSDFSQEPIMAL 2226
Qy 682 -----KCREMYDKDITIMLOIGASLMDPNKFLLLVLQVRYELAEAFNK 723
Db 2227 RTVLEILMEKEMDMSQRECIKDILTKHLVLSI-----LARTFKN 2267
Qy 724 T-ISTKDDDLIKOYNLT-----IEEMLQVLIYIVGERYVPGVGNVTKKEEVTMRBIIH 774
Db 2268 TOLPERAIFQIKQYNSVSGVSEWQLEE-AQVF-----WAKKEQSALSLIL 2312
Qy 775 LLCIEPMPHSATAKN-----LPENENNETGLENVINKVATFFKPGVSGHGCVY - 821
Db 2313 KOMIKKLDASCAANNPSLKLTYTECLRVCGNWLAECTLENPAVIMQTYLEKAVEVAGNYD 2372
Qy 822 -----ELKDESLKDFNMYYHYHYSKTOHSAEHMQK-----KRRKQE----- 857
Db 2373 GESSDELNRGKMKAF--LSLARFSDTOYQRIENYMKSEFENKQALLKRAKEEYGLRHEK 2431
Qy 858 -----NKDE-----ALPPPPPEFCFAPAFSKVINLLNCDIM-----MYILRTV 894
Db 2432 IQTRNYTVKQVQRELELDELALRALKEDRKFLCKA---VENYINCLLSGEGHDMWVRL- 2487

QY	895	FERAIDTDSNIWTEGMLQWAPHIALGLLEEKQOLQKAPEEVTFDF----	YHKASRLGS	950
Db	2488	-----CSLWLENS-----GYSEYNGMMKRGDKMKTPTKPLPLMYQLAARMGT	2529	
QY	951	SAMNIQMLLEKLGKIPQLEGOKDMITWILQMFDTVKRLREKSCULIVATTSGSESINDEI	1010	
Db	2530	KMMG-----GLGFHEVLNLLISRI-----SMDHPHHTFTIILALANAN-----RDEF	2571	
QY	1011	THDKAEAKRR-----KAEAA-----RLHROKIMAQMSALQKNFI----	1045	
Db	2572	LYKPEVARSRLITKNVPKQSSOLDEDRTEAANRLIICIRSRPQWRSVEALCDAVILA	2631	
QY	1046	-----ETHKLMYDNTSEMPCKEDSIMEESTPAV-----SDYS-----R	1079	
Db	2632	NLDATQWKTKRGINIPADQPITKLKNEIDVVPVMTKEIKVDHTGEYGNLVTIQSKRAEFR	2691	
QY	1080	IALGPKRGSPVTEKEVLTCI-----LCOEQEYVKIEN	1111	
Db	2692	LAGG-----VNLPLKIDCVGSDGKERQLVKRGDLRQDAVMQOVQMCNTLLQRNTET	2745	
QY	1112	NAMVLISACVOKSTALTOHRG-----KPIELSGEALDPLFMDPDPLAYGTYYTGCCHVM	1163	
Db	2746	RKRKLITICTYKVPVLSQRSGVLEWCTGVPPI---GEF---LVNNEDGAHKRYRPNDFSAF	2799	
QY	1164	HAVCWOKEPEAVOLSSOORIHVDLFLESGEVLCPCKSCINCTVIPIPLQPKINSENA	1223	
Db	2800	Q---CQKKMEVOKKSFEEKYEVFM-----DVCQNFQPVF-----RY	2833	
QY	1224	DALAOULLTLARMIQTVLA-----RISGYNIRHAKGNPIPIFFNQMGMDSTLBFHSI	1375	
Db	2834	FCMEKFLDPAIWFEXRLATRSVATSSIVGIL-----GLGDRHQVQ---2874		
QY	1276	LSRGVSSSTKYSNTEKMWILFATTIYRIGLUKVPDDRDP-----RVPMLT	1321	
Db	2875	-----NILINEQSAELVHIDGLGVAFBQGIILPTPTVFRLTRDIVDGMGITGVGEVF	2927	
QY	1322	WSTCAFTIOAIEN-----LLGDGKPLGALQNRQHNGCLKALMQFAVAORTCP	1370	
Db	2928	RRCCEKTEVMNRNSOETLLTIVVELLYD---PLF---DWTMNPALKALY---LQOR---P	2974	
QY	1371	QVLIQKHLVRLLSVVLPNPKISEDPTCLLSIDLFLHVLGAVLAPSLXWDDPVDLPQSSVS	1430	
Db	2975	EDETELUH-----PTLNADDQCECKNUL-----SDID	2999	
QY	1431	SSYNHLYLFHLITMAHMLQIILLTVDITGLPLA---QVQ---EDSEAHSSASSFAEISQYT	1484	
Db	3000	QSFDKVAERVLMLRQLEKUK---GVEEGTVLSGVGOVNLIIQQAIDPKNLSRLF-----	3049	
QY	1485	SGSIGCDIPGWYLW	1499	
Db	3050	-----PGWKAWV	3056	

RESULT 12

```

RESOLUTION 12
US-80-874-266-2
; Sequence 2, Application US/08874266
; Patent No. 5955279
; GENERAL INFORMATION:
; APPLICANT: Gatti, Richard A.
; TITLE OF INVENTION: ATAXIA-TELANGIECTASIA: MUTATIONS IN THE ATM GENE
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson and Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: USA
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS

```



```

; TELEFAX: 810-539-5055
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3056 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; US-08-952-014C-3

Query Match      1.4%; Score 133.5; DB 4; Length 3056;
Best local Similarity 17.9%; Pred. No. 0.061;
Matches 260; Conservative 196; Mismatches 470; Indels 529; Gaps

QY 319 SWNNKIMSYSSDFRIFCOACLR-----EPDSENP-----CL--ISRMLMWD 359
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 1857 SWRNLSHIV---GGFTSCLRHSQSRTSTPANLDSESHFFCCDDKKKSQRTMLAV 1912
    : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
QY 360 AKLYKGARKILHELIFS-SFFMEMEYKKL-----FAMEFYKYIKQLQKEYISDD 407
    : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 1913 VDMRRQRKPSSGTFINDAFWLDNLNLEVAKVAQSCAAHPTALLYAEIV--ADKSNDDQ 1970
    : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
QY 408 HDRSIITALSVMQTVTPTLARHLTEEQNVISVIHTLELVPEYLDNRNKNFNQGYSD 467
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 1971 EKRSUAFEEGS-QSTTISLSKSKBEETGI--SLQDLLLEIYRSIGEPSDSLYCGG--- 2023
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
QY 468 KLGRYVAVICDLKYILISKPTTWTERLMQLEGRFSFLKIITCMGMEEIRROVGQ-HI 526
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 2024 --GKMLOPITRLR--TYEHAMWGKALVTYDLE--TAIPSTRQAIIQAQNGLGCHI 2076
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
QY 527 EVDPDWEAIAIQMKLNILLMFQEWCADEELLVA-----YKECHKAVMRC-STSF I 579
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 2077 -----LSVYLKGLDYENKWDCPEELHLYQAARNMQWDHCTSVSKVEGTSYH 2125
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
QY 580 SSSKTVVOSCG-----HSLETYSRYVSDELVSITHPLSRTLAGLHVRLSRGAVSRL 631
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 2126 ESLYNALQSLRDREFTFYESLUYARVKEVEEMCKRSL---ESVSYLPTLSRLQAIGEL 2182
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
QY 632 HEFVSFEFOV-----EVLYEYPLRCLVLVAQVAEMWRNNGLSLSISOVFYYQDV--- 681
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 2183 ESIGELFSRVTHROLSEVYIK-----WQHSOLLKDSDFSQEPIMAL 2226
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
QY 682 -----KCREMYDKDIIMIQIGASLMDPNPKFLLLVLYORYELAEAFNK 723
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 2227 RTVILEILMEKENMSQRECINKDLTKHLVELSI-----LARTFKN 2267
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
QY 724 T-ISRKQODLIKQYNTL-----IEEMLOVLIYIVGERYPGVGNVKKEEVTMRIIH 774
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 2268 TOLPRATFIQIKOYNSVSGVSEWOLEE-AQVP-----WAKKEQSLALSIL 2312
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
QY 775 LLCIEPMPHSAIAKN-----LPENENNNETGLEVINKVATFKPKPGVSGHGVI- 821
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 2313 KOMIKKLDAASCANNPSLKLTYTECLRCVGCNWLAETCLENPVIMOTYLEKA VEVAGNYD 2372
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
QY 822 -----ELKDSELKDPNMVPHYHYSKTOHSAKHMQK-----KRRKQE----- 857
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 2373 GESSDELNRGKMKAF-LSLAREFSQTQYRIENYMKSSFEFENKOALLKRAKEVEGLLREHK 2431
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
QY 858 -----NKDE-----ALPPPPPPECAPASKVINLLNCDIM-----MYILRTV 894
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 2432 IOTNRYTVKVORELELDELALRALKEDRKRFLCKA---VENYINCULSGEEHDMWVFL- 2487
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
QY 895 FERAITDOSNLWTGEMLQMAFHIALGLLLEEQQQLKAPEEEVTFDF----YHKASRLGS 950
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 2488 -----CSLWLENS-----GVSEVNGMMKRDGMKIPTYKFPLPMYQLAARMGT 2529
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
QY 951 SAMNTQMILLEKUKGPOLPEGQKDMTWTWILOMDETVRRLREKSLCI VATTSGGSESIAKDEI 1010
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 2530 KNMG-----GLGFHEVLNLLISRI-----SMDPHPTHTFIILALANAN----RDEF 2571
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :

```

QY	1011	THDKKAEKRK	-----KAEEA-----RULHROIKMAQSALOKNFI----	1045
Db	2572	LTKPEVARRSITKNVPKQSSQLDEDRTEAANRIICTIRSRPQVYRSVEALCDAYIIILA	2631	
QY	1046	-----ETHKLMDNTSEMPGKEDSIMEESTPAV-----SDYS-----R	1079	
Db	2632	NLDATOWKTQRKGINIPADQPITHKLMLNEDVVPTMEIKVDHTGEYNLVTQSFKAER	2691	
QY	1080	IHALGPKRGPSYTEKEVLTCI-----LCQESEEQVKIEN	1111	
Db	2692	LAGG-----VNLPKIIDCVSGDGKERRQLVKGRDRLRDQAVMQVFQMCTLLOLRNET	2745	
QY	1112	NAMVLSACVQKSTALTQHRC-----KDPIELSGEALDPFLMDDPLDAYGTGYTGCSCCHVM	1163	
Db	2746	RKRKITICYKVWPLSQRSRGYLEWCCTGVPI---GEF---LVNNEDGAHKRYRPNDFSAF	2799	
QY	1164	HAVCMQKYFEAVOLVSQORIHVDLFDESGEYLCPCLKSCLNTVIPILPLOPKINSENA	1223	
Db	2800	Q---CQKKMWEVQKSFBEKYEVEFM-----DYCQNFPQVF-----RY	2833	
QY	1224	DALAOLLFLARWIOTVIA-----RISQYNIRHAKGENPIPFNQMGDSSTLEFHSI	1275	
Db	2834	FCMEXFLDPAIWFEKRLAYTRSVATSSIVGIL-----GLGDHRHQ-----	2874	
QY	1276	LSFGVESSIKYSNSTIKEMVILFATTIYRIGLKVPPODERP-----RVPMILT	1321	
Db	2875	-----NILNEOGAELVHIDLGVAFEQGKIILTPETVPFRITRDIVDMGTGVEGVF	2927	
QY	1322	WSTCAFTQAIEIEN-----LLGDEGKPLFGALQNRHGNLKMALFOFAVAORITCP	1370	
Db	2928	RRCCEKTMVMRNSQETLLTIVEVLLYD---PLF---DWTMNPALKALY---LQQR---P	2974	
QY	1371	QVLIQKHVIRLLSVVLPNIKSEDTPCLLSIDLPHVLVGAVLATFPSLYWDDPDVLDPPSSVS	1430	
Db	2975	EDETELH-----PTLNADDOECKRNL-----SDID	2999	
QY	1431	SSYNHLYLPHLITMAHMLQILLTVDTGLPLA---QVO---EDSEEAHSASSFFAEISQYT	1484	
Db	3000	QSFDKVAERVLMRLOEKLIK---GVEECTVLSVGQYNLLIQOAIIDPKNLSRLF-----	3049	
QY	1485	SGSIGCDIPGWYLMW	1499	
Db	3050	-----PGWKAWW	3056	

Search completed: February 28, 2002, 10:02:00
Job time: 297 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 28, 2002, 09:56:13 ; Search time 35.89 Seconds
(without alignments)
3609.757 Million cell updates/sec

Title: US-09-724-126A-2
Perfect score: 9224
Sequence: 1 MADEPAGGTERMEISAEPLQ.....ETIARQETNOMLFGFNWQLL 1749

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_ll01.*
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8617	93.4	1757	AAW84351	Murine ubiquitin-p
2	8617	93.4	1757	AAW84351	Amino acid sequenc
3	4209	45.6	811	AAW84351	Human protein sequ
4	2710.5	29.4	1109	AAW84351	Polypeptide identi
5	1716	18.6	333	AAW84351	Partial human ubiq
6	1716	18.6	333	AAW84351	Amino acid sequenc
7	296	3.2	250	AAW84351	Human protein sequ
8	295	3.2	247	AAW84351	Human secreted pro
9	197.5	2.1	2096	AAW84351	Human ORFX ORF1356
10	189.5	2.1	2048	AAW84351	Human polypeptide
11	184	2.0	1411	AAW84351	Nucleolar/endosoma

12	182	2.0	2442	21	AAW77575	Human cytoskeletal
13	176	1.9	2688	22	AAW40883	Human polypeptide
14	175.5	1.9	1427	12	AAW10534	Human 160KD mediat
15	175.5	1.9	2663	22	AAW39097	Human polypeptide
16	174.5	1.9	1392	20	AAW09999	Restin protein seq
17	173	1.9	2301	22	AAU14720	Novel bone marrow
18	172.5	1.9	1780	22	AAW38681	Human polypeptide
19	172	1.9	1788	22	AAW40467	Human polypeptide
20	170	1.8	2298	22	AAU14794	Novel bone marrow
21	164.5	1.8	2139	22	AAW47278	PN7771 Homo sapi
22	164	1.8	2954	20	AAW01632	Amino acid sequenc
23	156.5	1.7	2710	17	AAW95016	C. difficile toxin
24	156.5	1.7	2710	19	AAW68387	Clostridium diffic
25	155	1.7	3685	10	AAW90290	Human Duchenne mus
26	154	1.7	1562	21	AAW39233	Arabidopsis thalia
27	154	1.7	3685	10	AAW90373	Sequence encoded b
28	152.5	1.7	980	21	AAW18294	Plasmodium faicipa
29	152.5	1.7	1325	18	AAW19540	Male-enhanced anti
30	152.5	1.7	1325	20	AAW94391	Mouse male enhance
31	151	1.6	1521	21	AAW39235	Arabidopsis thalia
32	151	1.6	1528	21	AAW39234	Arabidopsis thalia
33	151	1.6	1703	21	AAW36714	Arabidopsis thalia
34	151	1.6	1710	21	AAW36713	Arabidopsis thalia
35	151	1.6	1744	21	AAW36712	Arabidopsis thalia
36	151	1.6	5447	22	AAU14697	Novel bone marrow
37	150	1.6	5373	22	AAU14603	Novel bone marrow
38	149	1.6	3433	18	AAW22017	Utrophin. Homo sa
39	148	1.6	3021	18	AAW19661	ATM mutant 6348del
40	145.5	1.6	677	20	AAW85608	Secreted protein c
41	145.5	1.6	677	22	AAW64560	Human myosin heavy
42	144	1.6	3003	18	AAW19658	ATM mutant 7630del
43	143.5	1.6	3262	21	AAW53675	Mechanical stress
44	143	1.6	1150	22	AAU14626	Novel bone marrow
45	142.5	1.5	1398	21	AAW18292	Plasmodium faicipa

ALIGNMENTS

RESULT 1

AAW84351
ID AAW84351 standard; Protein; 1757 AA.

AC AAW84351;

DT 25-MAR-1999 (first entry)

DE Murine ubiquitin-protein ligase Ubql.

UBiquitin-protein ligase; Ubql; mouse; ubiquitinylation; degradation;
N-end rule pathway; stress-related muscle wasting; inhibitor; screen.

OS Mus sp.

PN US5861312-A.

PD 19-JAN-1999.

PF 02-DEC-1997; 97US-0982956.

PR 02-DEC-1997; 97US-0982956.

PA (CALY) CALIFORNIA INST OF TECHNOLOGY.

PI Kwon YT, Varshavsky A;

DR WPI: 1999-130395/11.

DR N-PSDB; AAW99308.

PT Mouse and human Ubql cDNA - useful for producing recombinant Ubql

XX polypeptides

PS Disclosure; Columns 15-28; 18pp; English.

```

XX The present sequence represents a ubiquitin-protein ligase called Ubr1.
CC The Ubr1 enzymes are involved in protein ubiquitinylation and
CC ultimate degradation through the N-end rule pathway and have been
CC linked to stress-related muscle wasting. Recombinant Ubr1 polypeptides
CC can be used to screen for inhibitors of muscle wasting when this is
CC associated with the N-end rule pathway.
XX
SQ Sequence 1757 AA;

Query Match 93.4%; Score 861.7; DB 20; Length 1757;
Best Local Similarity 92.1%, Pred. No. 0;
Matches 1619; Conservative 64; Mismatches 66; Indels 8; Gaps 3;

QY 1 MADEEAGGTMEIEISAEPLQTPORLASWMDQVDFYAFHLHLAQLVPEIYFAEMDPDLE 60
DB 1 madeemdaermdvpeplapqrpaswddqvdftafhlhlaqlvpeiyfaemdpdle 60

QY 61 KOESVQMSIFTPLEWYLEGEDPDICLEKLKHSAGFQLCGRVFKSGETTYSCRDCAIDPT 120
DB 1 koesvqmsilltplewylfgedpdiclekklkhsagfqlcgkvfksgettyscrdcaidpt 120

QY 121 CVLWDCFDOSVHKHRYKWHSTGCGFCDCGDTFAWKTPGFCVNHPEPRACTIKENSRC 180
DB 1 cvlwmcdcfcgssvkhkrykwhstgggfcgdcgdtfaektpgfcvnhpepactikensrc 180

QY 181 PLNEEVIVQARKIFPSVIKYVEMTIWEPEKLPBPLOIRKERNERYCVLFNDEHSHSYDH 240
DB 1 plneeviaqarrifpsvkiyvemtiweekeklpelploirkernerycvlfndehshsydh 240

QY 241 VYISQORALDCELAEAQLHTTAIDKEGRAVYKAGAYAAQEAKEKDKSHSENVVQHPHV 300
DB 241 vysiqraldcelaeqlhttaidkegravkagayaaqeaekedkshsenvsqphlv 300

QY 301 EYVLSHSEIMAHQFAIRLGSWMNKINSYSSDFRQIFCOACLRPEPDSFNCLLSRLMWD 360
DB 301 evlshseimahqfairlgswnkninsysddfrrqifcoaclreepdsfncllsrlmwd 360

QY 361 KLYKGARKILHELIFSSPFMEYKFLFAMEFYKYYKQLOKEVYISDDHRSISITALSQV 420
DB 361 klykgarkilhelifsspfmeneykklfamefykyykqlqkevysddhersiitalsvq 420

QY 421 MFTVPTLARHLIEQNVISVIFETLEVLPEYLDNRNKNFPGYSODKILGRVYAVTCDLK 480
DB 421 mftvptlarhlieeqnvisevitellevpeyldrnknfnfgysqdklgrvyavtcdlk 480

QY 481 YILISKPTTWTBELRMQFLEGRFSFLKILTCMQGHEIIRQVGOHIEVDPDWEAAIAQM 540
DB 481 yiliskpvtwtelrmqflegfrsflkiltcmqgmeeirrvqghievdpdweaaiaqm 540

QY 541 QLNKNTLLMFOEWCACDEELLVAYKECHKAVMRCSTSFSSKTVVQSCGSHLETYSYRV 600
DB 541 qlnkntllmfoewcacdeellvaykechkavmrcstsfsskttvvqscgshletksyrv 600

QY 601 SEDLVSIHLPLSRTLGLVRLSLRSLRSLRSLRSLRSLRSLRSLRSLRSLRSLRSLR 660
DB 601 sedlvsihlplslrtlaghlvrlslrslrslrslrslrslrslrslrslrslrslrsl 660

QY 661 AEMWRNRLGSLISQVFPYODVCKREMYDKDIIMLOIGASLMDPNKFLLLVLQRYELAE 720
DB 661 aemwrnrlgslisqvfpdyodvckremydkdiimloigaslmdpnkflllvlqryelta 720

QY 721 FNKTTSTKQDQDLIKOYNTLIEMLQVLIIVGERYVPGVGNVTKERTVREITHLLCIEP 780
DB 721 fnkttstkddqdlkqyntlieemlqvliivgeryvpgvgnvtnkervtreithllciep 780

QY 781 MPBSAIKALNPENENNETGLENVINKVAFKPKPGVSGHGVYELKDESLKDFNNFYHYSK 840
DB 781 mpbsaiarlnpenennetglenvinkvatfkkpgvsgghgvlyelkdeslkdfnmfyhyk 840

QY 841 TQHSKAEHMQKRRKQENKDEALPPPPPPFCFAFSKVINLLNCDIMMYILRTVFERAID 900
DB 841 tqhskaehmqkrrkqenkdcalppppppfcfafsksvinllncdimmyilrtvferaid 900

```

```

DB 841 tqhskaehmqkrrkqenkdcalppppppfcfafsksvinllncdimmyilrtvferaid 900
QY 901 TDSNLWTEGMLQMAFHILALGLEEKQOLQKAPEEVEVDFYHKASRLGSSAM--NIQM 957
DB 901 tesnlwtgmlqmafhiilalgleekqlkapeeevafdfyhkasrlgssamnaqlqm 960
QY 958 LLEKLKGIPQLEGOKDMITWILQMPDVTVKRLREKSKLIVATTSGSESTKNDLTHDKKA 1017
DB 958 llerklkgipqlgqgkdmittwilmqdmftvkrirreksclvattsgleciseethdkeka 1020
QY 1018 ERKKAARLHROKIMAOVSALOKNFTETHTKLMYDNTSEMPGKEDSIMEESTPAVS DY 1077
DB 1021 erkkacaarlhrdkimagsalqknfiethklmydntsevtgkedsimeestavsea 1080
QY 1078 SRTALGPKRGPSVTEKEVLTICLQEQEVKIENNAMVLSACVQKSTALTQHRGPIELS 1137
DB 1081 srtalgpkrpavtekeveltliclqeeqevklennamvlsacvqkstaltqhrgpvdhl 1140
QY 1138 GEALDPLFMDPDLAYGYTGTSCGHVHMAVCWQKYFEAVOLSSQQRHVDLFDLESGEYLC 1197
DB 1141 getldplfmdpdlaygtgtscghvmhvacwqkyfeavqlssqqrhvdldflesgeylc 1200
QY 1198 PLCKSLCNTVPIIPLQPKINSENADALAOILLTARWITQTVLARIISGYNIRHAKGENP- 1256
DB 1201 plckslcntvpiilpqpkinsenaalqiltarwigtvlarisgynirhakeapa 1260
QY 1257 IPIPFNMGDSTLEFHSILSGVSESSIKYSNIKEMWILFATTIYRIGLVKVPPODERPR 1316
DB 1261 vpvlfngmgdstlefhsilsgvssvksynsikemwiltattiyriglvkvppeidpr 1320
QY 1317 VPMLTWSTCAPTIQAIENLLGDEGKPLFGALQNRHNGLKALMQFAVAQRITCPQVLIQK 1376
DB 1321 vpmmtwstcatfiqaienllgdegkplfgalqnrhnglkalmqfavaqratcpqvlhik 1380
QY 1377 HLVLRLSVLPNIKSEDPCLLSLDLPHVLGAVLAFPSLYWDDPVDLQPSVSSSYNHL 1436
DB 1381 hlarllsvlpniksentscpdllsvdlphvlvgavlafpslywddcvdlqpslssynhl 1440
QY 1437 YLFHLITMAHMLQILLTVDT---GLPLAQVQEDSEEAHSASSFFAEISQYTSSTGCDI 1492
DB 1441 ylfhlitmahmlqillttdtldspgplaegeedseearcasaffvevsghtdgltcgca 1500
QY 1493 PGWYLVSLKNGITPYLRCAALFPHYLLGVTPPELHTNSAEGEYSALCSYLSLPTNLPL 1552
DB 1501 pgwylvslsrlngitpylrcaallfhyllgvappeelfansaegetsalcsylslptnlfl 1560
QY 1553 LFOEYWDTVREPLQRCWADPALLNCLKOKNTVVRYPKRNLSLIELDDYSCLLNOASHPR 1612
DB 1561 lfgeywdttriplqrwcgdpallkslkqksavrvyrprknslielpedyscllnqashfr 1620
QY 1613 CPRSADDERKHPVLCLFCGAILCSNICCQEIIVNGEEVGACITFHALHCGAGVCIPLKIRE 1672
DB 1621 cprsadderkhvplclfcgailcsnicccqeilvnggeevgacvfhahcgagvciflkire 1680
QY 1673 CRVVLVEKGARGCAYPAPYLDYGETDPGLKRGPNPLHLSRERYRKLHLVWQOHCITEETA 1732
DB 1681 crvvlvkgargcaypapyldygetdpglkrpnplhlsreryrklhlvwqghciieela 1740
QY 1733 RSQETNOMLFGFNWQLL 1749
DB 1741 rsqetnmlfgfnwql 1757

RESULT 2
AAB31162
ID AAB31162 standard; Protein; 1757 AA.
XX
AC AAB31162;
XX
DT 02-APR-2001 (first entry)
XX
DE Amino acid sequence of Mouse Ubr1 protein.
XX

```


KW Ubql; E3-type protein; ubiquitin system; ubiquitin-protein ligase;
 KW N-end rule pathway; intracellular pathogen; Lysteria monocytogenes;
 KW Yersinia enterocolitica; muscle wasting; infection.

OS Mus sp.

XX US6159732-A.

XX 12-DEC-2000.

XX 11-JAN-1999; 99US-0228317.

XX 02-DEC-1997; 97US-0982956.

XX (CALY) CALIFORNIA INST OF TECHNOLOGY.

XX Kwon YT, Varshavsky A;

XX WPI; 2001-090278/10.

DR N-PSDB; AAC86933.

XX Inhibiting the N-end rule pathway in mammalian cells for treating
 PT infections and various diseases associated with muscle tissue wasting,
 PT by inhibiting the expression of Ubql gene

XX Example; Column 15-28; 18pp; English.

XX The present sequence represents a murine Ubql enzyme. Ubql is an E3-type
 CC protein of the ubiquitin system. Specifically, it is a ubiquitin-protein
 CC ligase. The enzyme is specific for destabilising residues exposed at
 CC the N-terminus of protein substrates. Inhibition of the expression of
 CC Ubql gene in a cell results in inhibition of the N-end rule pathway.
 CC The method is used for treatment of mammalian cells infected with an
 CC intracellular pathogen, e.g. Lysteria monocytogenes or Yersinia
 CC enterocolitica. Inhibition of N-end rule pathway is also useful for
 CC treating various diseases associated with wasting of muscle tissue and
 CC infections.

XX Sequence 1757 AA;

Query Match 93.4%; Score 8617; DB 22; Length 1757;
 Best Local Similarity 92.1%; Pred. No. 0;
 Matches 1619; Conservative 64; Mismatches 66; Indels 8; Gaps 3;

Qy 1 MADEEAGTERMETISAEPLQTPQRLASWWDQVDFYTAFLHHLAQLVPEIYFAEMDPDL 60

Db 1 madeemgaermvdspepplapgrpaswddqvdftytaflhhlagvlpeiyfaemdpdl 60

Qy 61 KOESVQMSITPLEWYLFQGDPOICLEKLGKSHGAFOLCGRVFKSGETYSRCDCAIDPT 120

Db 61 kreesvqmsitplewylfgedpdaiclekshgafqlcgkvfsggetysrccdaidpt 120

Qy 121 CVLMDCFQDSVHKHRYKMHSTGGGFCDCGDTFANKTPFCVNHPEPRAGTICKENSRC 180

Db 121 cvlmdcfqsvhkhryknhstsggfcddgdtewktpfcvndhepragttckeslhc 180

Qy 181 PLNEEIVQARKFPSPVIVKVVEMTIEEKEPPELQIREKNERYCVLENDEHSHSDH 240

Db 181 plneevlaqarrifpsvivyemtiweeeekelpelqireknerycvlendehshsydh 240

Qy 241 VIYSLQALDCELAELAAOLHTTAIDKEGRRAVKAGAYAACQEAKEEDIKSHSENVSOHPLHV 300

Db 241 viyslqalceelaqaqlhttaldegravravagvatccgeakedikshsenvsqphlv 300

Qy 301 EVLHSEIMAHQKALRLGSGWMNKIMSYSSDFRQIFCQACLRPEEDSENPCILISRLMLWDA 360

Db 301 evlshvmahqkfalrlgsgwmnkimsysdfrqifcqaclveepgsenpcilislmlwda 360

Qy 361 KLYGARKILHELIFSSFFMEMEYKKLFAMFVYKQLQKEYISDDHDSRISITALSVQ 420

Db 361 klygarkilhelilfssffmemeykklfamefvkykqlqkeyisddhdersitalsvq 420

Qy 421 MFTVPTLARHLIEEQNVISVITETLLLEVLPEYLDNRNKNFQGYSDQKLGKRVYAVICDLK 480

Db 421 mltvptlarhlieeqnvsvitelllevlpeylidnrnknfngqysdqkldgrvyavicdlk 480

Qy 481 YLILSKPTTWTERLRMQFLGFRSFLKILTCMQGMEIRQVQGHIEVDPPWEAIAIQM 540

Db 481 ylliskpvtwterlrqflgfrsflkiltcmqgmeeirrvqvgqhlievdpdweaiaiqm 540

Qy 541 QLKNIILMFQEWACACDEELLVAVKECHKAVMRCSFISSTKVTVOSCGHSLSTKSYRV 600

Db 541 qlkniilmfqcacdedlllvaykechkavmrctsmfctkvqlcglnslctksykv 600

Qy 601 SEDLSIHPLSRTLAGLHVRLSRLGAVSRLEHFVSFEDFQVEVLVPEYRCLVLVAQV 660

Db 601 sedlsvihplsrtaglhvrlslrgaisrlhefvpfidsqvevlvepvlrclvlvaqv 660

Qy 661 AEMWRRNGLSLSQVFFYQDVVKREEMWDKDIIMLQIGASLMDPNKFLLLVLQRYEAE 720

Db 661 aemwrrnglslsqvfyqdvkcreemydkdiimlqigaslmdpnkflillvlqryeld 720

Qy 721 ENKTISTKDDLLKOYNTLIEEMLOVLIYVGERYVPGVGNVTKKEVTMREIHLCLIEP 780

Db 721 enktistkdqllikgnyntlieemlvliyvgerypvgvgnvtreevimreithlcliep 780

Qy 781 MPHSATAKNLPENNNETGLENVINKVATPKPGVSGHGVYELKDESLKDFNMFYHYSK 840

Db 781 mpsaiarlpenennetglenvinkvatfkpgvsgbhgyelkdeslkdfnmfyhysk 840

Qy 841 TOHSAEAMOKRRKQENKDEALPPPPPECFPAFSKVINLLNCDIMMYLRTVFERAID 900

Db 841 tqhsaehmqkrrkqenkdalpppppefcfapfskvvlnllscdmvlylrltiferavd 900

Qy 901 TDSNLWTEGMLQAFHLALGALLEEKQLOKAPEEYTFDFYHKASRLGSSAM--NLQM 957

Db 901 tesnlwtgmlqafhlalalgleekqldkapeeevafdfyhkarslgssamaqlgm 960

Qy 958 LLEKLKGIQLEQOKDMITWILQMFDTVKRLREKSLCIVATTTSSESTKNDIEITHDEKA 1017

Db 961 llerlkgiqlegkdmiltwilmfdtvlrklreksclivatttsgelecieeithdkeka 1020

Qy 1018 ERKRKAARLHRQKIMAQMSALOKNFIETHKLMYDNTSEMPGKEDSIMPEESTPAVS DY 1077

Db 1021 erkrkaearlhrqkimaqmsalqknfiethklydntsevtgkedsimeeestsavsea 1080

Qy 1078 SRLALGPKGPSYTEKEVLTCLICQEBQEVYKIENNAVLNACVOKSTALQHRCKPIELS 1137

Db 1081 srlalgpkrpavtekevtclilcqeegvklennamvlsacvdkstaltqhrckpvdhl 1140

Qy 1138 GEALDPLFMDPLAYGTYTGSCGHVMHVCWKYFEAVQLSSQORIHVDLFDLESGEYLC 1197

Db 1141 getldplfmdplahgtytgscghvmhvcwkqyfeavqlssqqrhvdldflesgeylc 1200

Qy 1198 PLCKSLCNTVPIPILOPQKINSADALAOQLTLARMIQTVLARIISGYNIRHAKGENP- 1256

Db 1201 plcksicntvpiplpqpkinsenacalaqltlarwiqtvlariisgynikhakeapa 1260

Qy 1257 IPFIFFNOGMDSTLEFHSILSFGVSESIKYSNISKEMVILPATIYRIGLKVPPDERDPR 1316

Db 1261 vplffngmgdstfshsilsgvsvksysnikemvilfattiyriglkvpvdeidpr 1320

Qy 1317 VPMLTWSTCAFTTQATENLLGDEGKPLFGALQNRHNGKALMOFAVAORITCPOVLIOK 1376

Db 1321 vpmmtwtstcafttqatlenllgdegkplfgalnqrhsglkalmafavaqrattcpqvlilh 1380

Qy 1377 HLVRILSVVLNPINIKEDTPLLIDLFHVLVGAFLAEPSPYWDVDPDPSPSSSSSYNHL 1436

Db 1381 hlarilsvilpnlsqentpgllsvdlfhvlvgavlaflpslywdtdvdlqpsplssynhl 1440

Qy 1437 YLFHLITMAHMLQILLTVDT-GLPLAQVEDSESAHSAASSFAEISQVTSIGCGDI 1492

Db 1441 ylfhlitmahmlqillttdtldspgplaegeedseearcasatfvevsgtdlgtcgca 1500

Qy 1493 PGWYLVSLKNGITPYLRCAALFFHYLLGYTPPELHTNSAEGEYSALCSLSLPTNLFL 1552

Db	1501	pwylwlsirngitpylrcaallfnylglvapeelfansaegelsaicsylslptnlfl	1560
QY	1553	lfqeywdtivrplliqwcadpallncLKOKNTVVRYPKRKNSLIELPDDYSCLLNQASHFR	1612
Db	1561	lfqeywdtirlplqwcgdpaillkslkqksavvyrprknslielipedyscillnqashfr	1620
QY	1613	cpRSADDERKHPVLCIFCGAILCSNICQEIIVGEEVGACIFHALHCGAGVCIFLKIRE	1672
Db	1621	cpRSadderKhpVLCIfcgailcsqnicqeiivngeevgacvfhalhcgagvciflkire	1680
QY	1673	CRVWLVEGKARCAYPAPVLDYGETDPClKRGNPILHSRERYKHLHWQOHCIIIEETA	1732
Db	1681	crvvlvegkarcaypapyldygetdpglkrgnplhlsrerykrhlhwgqhciiieeta	1740
QY	1733	RSQETNOMLFGFNWOLL 1749	
Db	1741	rsqetnqmlfgfnwql 1757	
RESULT 3			
AAB93464			
ID	AAB93464 standard; Protein; 811 AA.		
XX	AAB93464;		
XX	26-JUN-2001 (first entry)		
XX	Human protein sequence SEQ ID NO:12732.		
DE	Human; primer: detection; diagnosis; antisense therapy; gene therapy.		
XX	Homo sapiens.		
KW	EP1074617-A2.		
XX	07-FEB-2001.		
PD	28-JUL-2000; 2000EP-0116126.		
XX	29-JUL-1999; 99JP-0248036.		
PR	27-AUG-1999; 99JP-0300253.		
PR	11-JAN-2000; 2000JP-0118776.		
PR	02-MAY-2000; 2000JP-0183767.		
PR	09-JUN-2000; 2000JP-0241899.		
XX	(HELI-) HELIX RES INST.		
XX	Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;		
PI	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;		
XX	WPI: 2001-318749/34.		
DR	Primer sets for synthesizing polynucleotides, particularly the 5602		
PT	full-length cDNAs defined in the specification, and for the detection		
PT	and/or diagnosis of the abnormality of the proteins encoded by the		
PT	full-length cDNAs -		
XX	Claim 8; SEQ ID 12732; 2537pp + CD ROM; English.		
PS	The present invention describes primer sets for synthesizing 5602		
XX	full-length cDNAs defined in the specification. Where a primer set		
CC	comprises: (a) an oligo-dT primer and an oligonucleotide complementary		
CC	to the complementary strand of a polynucleotide which comprises one of		
CC	the 5602 nucleotide sequences defined in the specification, where the		
CC	oligonucleotide comprises at least 15 nucleotides; or (b) a combination		
CC	of an oligonucleotide comprising a sequence complementary to the		
CC	complementary strand of a polynucleotide which comprises a 5'-end		
CC	sequence and an oligonucleotide comprising a sequence complementary to a		
CC	polynucleotide which comprises a 3'-end sequence, where the		
CC	oligonucleotide comprises at least 15 nucleotides and the combination of		
CC	the 5'-end sequence/3'-end sequence is selected from those defined in		
CC	the specification. The primer sets can be used in antisense therapy and		

	in gene therapy. The primers are useful for synthesising polynucleotides particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.
xx	
SQ	Sequence 811 AA;
	Query Match 45.6%; Score 4209; DB 22; Length 811; Best Local Similarity 100.0%; Pred. No. 0; Matches 811; Conservative 0; Mismatches 0; Indels 0; Gaps
QY	204 MTIMEEKEPELOIREKNERYCYVLFENDEHHSYDHVIYSIORALDCELAEAOLHTAI 263
DB	1 mtiweeekeipbelgireknerycyvlfndehhsydhviyslgralcelaeaahhtal 60
QY	264 DKERRAVKAGAYAAQCQAEAKEDI KSHSENVSQHPLHVLEVLHSEITMAHQEKALRLGSMWK 323
DB	61 dkerravkagayaaqcqaekedikshsenvsqphlvlevlhseimabqhkalrlgsmwk 120
QY	324 IMSYSSDFRQLFCQACLRREPDPSENPCLSRLMLWDAKLKGARKILHELIFSSFME 383
DB	121 imsyssdfrcqlfcacclreepdsenpcclsrlmlwdaklykgarkilhelifssffme 180
QY	384 YKKLIFAMEFVKYKLOKEYI ISDDHDRSISITALSVQMFTVPTLARHLIEEQNVISITE 443
DB	181 ykklifamefvkykqlkeyisddhdarsitalsvgmftvptlarhlieeqnvisite 240
QY	444 TLLEVLPYLDRNNKFNGFGYSQDKLGRVYAVICDLKYILISKPTIWTERLMQFLGFR 503
DB	241 tllevlpylatnnkfngfsgdklgrvyavicdlkyllskptiwtelrlnmqflgfr 300
QY	504 SFLKILTTCMQMEEETRQVGQHI EVDPDWEAATAIQWLKNILLMPQEWCADEELLVA 563
DB	301 sflkiltcmqmeeetrvgghievdpdweataiaigmqlknillmfqewcacdeellva 360
QY	564 YKECHKAVMRCSSTFISSSKTVVQSCCHSLTETKSRYVEDLVSIHLPLSLTAGLHVRLS 623
DB	361 yechkavmrscstsfissktvvqscgchsltetsryvedlvsihlplstlaglhvrls 420
QY	624 RLGAVSRUHFVFEDFOVEYLVEYPLRCLLVLVAQVVAEMRRNGLSLSIOVFYYQDVKC 683
DB	421 rlgavsrlhevfesfdfeveylveypnlrlvlvaqvvaemrrnglsisqgfyyqvdkc 480
QY	684 REEMYDKDIIMLIQASLMDPNKFL LLVLQRYELAEAFNKTISTKOODLIKQNTLIEEM 743
DB	481 reemydkdiimligaslmnpnkfl llvglryelaefnfktistkdodlikqyntlieem 540
QY	744 LQVALIYIGERVPGVGNVTKEEVTMRIIHLLCIEPMPHSATAKNLPENNETGLENV 803
DB	541 lqvliyigerpypgvgnvtkeevtmrei ihllciepmphsaiaaknlpenennetglenv 600
QY	804 INKVATFKKPGVSGHVVELKD ESUKDFNMYPHYHYSKTHQSKAHEHMOKKRKQENKDEAL 863
DB	601 inkvatfkpgvsghvvelkdesl kdfnmypfhyyskthgskaehmokkrkrkenkdeal 660
QY	864 PPPPPPECFAFSKVINLLNCDIMMY ILRTVPFERADTDNSLNLTGECMLQNAFHILA GLLL 923
DB	661 ppppppecpafskvinllncdimmy ilrtvferadtdnsnlwtegmqlmafhlailgl 720
QY	924 BEKOOLQAPBEEVTFDFYHKASRLGSSAMNIOMLLEKLGIPLEGQCKDMITWILOMF 983
DB	721 eekoolqapeeevtfdfyhkasrl gssamniomlleklkiplegqckdmiltwilomfd 780
QY	984 TVKRLREKSLIVATTSSGESIKND EITHDK 1014
DB	781 tvkrlreksc livattssgesikndeithdk 811

QY	817	GHGVYELKDESLKDFNMIFYHYISQTSKHAHMQRKRQRKDEKDALPP
----	-----	---

DE Partial human ubiquitin-protein ligase, Ub1.
 XX Ubiquitin-protein ligase; Ub1; human; ubiquitinylation; degradation;
 KW N-end rule pathway; stress-related muscle wasting; inhibitor; screen.
 XX Homo sapiens.
 XX US5861312-A.
 XX 19-JAN-1999.
 XX 02-DEC-1997; 97US-0982956.
 XX 02-DEC-1997; 97US-0982956.
 XX (CALY) CALIFORNIA INST OF TECHNOLOGY.
 PA Kwon YT, Varshavsky A;
 PI WPI; 1999-130395/11.
 DR N-PSDB; AAX03300.
 XX Mouse and human Ub1 cDNA - useful for producing recombinant Ub1
 PT polypeptides
 XX Disclosure; Columns 27-30; 18pp; English.
 XX The present sequence represents a partial ubiquitin-protein ligase called
 CC Ub1. The Ub1 enzymes are involved in protein ubiquitinylation and
 CC ultimate degradation through the N-end rule pathway and have been
 CC linked to stress-related muscle wasting. Recombinant Ub1 polypeptides
 CC can be used to screen for inhibitors of muscle wasting when this is
 CC associated with the N-end rule pathway.
 XX Sequence 333 AA;
 SQ

Query Match 18.6%; Score 1716; DB 20; Length 333;
 Best Local Similarity 99.7%; Pred. No. 9.9e-137;
 Matches 332; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 702 MDPNFKLLVLQRYELAEAFNKTISTKQDQDLIKQYNTLIEEMLQVLIYVGERYPGVGN 761
 DB 1 mdpnkfllvlqryelaeafnktistkdqdlkqyntlieemlqvliyvgerypgvgn 60
 QY 762 VTKEEVTMRIIHLCTIEPMPSAIAKNLPENNETGLENVINKVATFKKPGVSGHGY 821
 DB 61 vtkeevtmreiihlctiepmphsaiaaknlpennnetglenvinkvatfkkpgvsgngvy 120
 QY 822 ELKDESCLKDFNMVYHYSKTOHSAEHMOKRRKQENKDEALPPPPPEFCFAFSKVINL 881
 DB 121 elkdeslkdfnmyfyhysktqhskaehmqkrrkqenkdalpppppfcfapfskvinl 180
 QY 882 LNCDDIMMYILRTVFERAIDTDSNLWTEGMLQMAFHILALGLLEEKQLOKAPEEVTFDF 941
 DB 181 lncddimmyilrtvferaintdsnlwtegmqlmafhilalglleekqqlkapeeevtfdf 240
 QY 942 YHKASRLGSSAMNIOMLLEKLGIPQLEGQKDMITWILQMDTVKRLKSKCLIVATTSG 1001
 DB 241 yhkasrlgssamniomlleklgipqlegqkdmittwilqmdftvkrlekscclivattsg 300
 QY 1002 SESIKNDEITHDKAEAKRRKAEAAARLHRQKIM 1034
 DB 301 sesikndeithdkeakerrkaeaaarlhrqkim 333

RESULT 6
 AAB31163
 ID AAB31163 standard; Protein; 333 AA.
 XX
 AC AAB31163;
 XX
 DT 02-APR-2001 (first entry)

XX Amino acid sequence of a partial human Ub1 protein.
 DE
 XX
 KW Ub1; E3-type protein; ubiquitin system; ubiquitin-protein ligase;
 KW N-end rule pathway; intracellular pathogen; Lysteria monocytogenes;
 KW Yersinia enterocolitica; muscle wasting; infection.
 XX
 OS Homo sapiens.
 XX
 XX US6159732-A.
 XX 12-DEC-2000.
 XX 11-JAN-1999; 99US-0228317.
 XX 02-DEC-1997; 97US-0982956.
 XX (CALY) CALIFORNIA INST OF TECHNOLOGY.
 PA Kwon YT, Varshavsky A;
 PI WPI; 2001-090278/10.
 DR N-PSDB; AAC86934.
 XX Inhibiting the N-end rule pathway in mammalian cells for treating
 PT infections and various diseases associated with muscle tissue wasting,
 PT by inhibiting the expression of Ub1 gene
 XX Example; Column 27-30; 18pp; English.
 XX The present sequence represents a partial Ub1 enzyme. Ub1 is an E3-type
 CC protein of the ubiquitin system. Specifically, it is a ubiquitin-protein
 CC ligase. The enzyme is specific for destabilising residues exposed at
 CC the N-terminus of protein substrates. Inhibition of the expression of
 CC Ub1 gene in a cell results in inhibition of the N-end rule pathway.
 CC The method is used for treatment of mammalian cells infected with an
 CC intracellular pathogen, e.g. Lysteria monocytogenes or Yersinia
 CC enterocolitica. Inhibition of N-end rule pathway is also useful for
 CC treating various diseases associated with wasting of muscle tissue and
 CC infections.
 XX Sequence 333 AA;
 SQ

Query Match 18.6%; Score 1716; DB 22; Length 333;
 Best Local Similarity 99.7%; Pred. No. 9.9e-137;
 Matches 332; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 702 MDPNFKLLVLQRYELAEAFNKTISTKQDQDLIKQYNTLIEEMLQVLIYVGERYPGVGN 761
 DB 1 mdpnkfllvlqryelaeafnktistkdqdlkqyntlieemlqvliyvgerypgvgn 60
 QY 762 VTKEEVTMRIIHLCTIEPMPSAIAKNLPENNETGLENVINKVATFKKPGVSGHGY 821
 DB 61 vtkeevtmreiihlctiepmphsaiaaknlpennnetglenvinkvatfkkpgvsgngvy 120
 QY 822 ELKDESCLKDFNMVYHYSKTOHSAEHMOKRRKQENKDEALPPPPPEFCFAFSKVINL 881
 DB 121 elkdeslkdfnmyfyhysktqhskaehmqkrrkqenkdalpppppfcfapfskvinl 180
 QY 882 LNCDDIMMYILRTVFERAIDTDSNLWTEGMLQMAFHILALGLLEEKQLOKAPEEVTFDF 941
 DB 181 lncddimmyilrtvferaintdsnlwtegmqlmafhilalglleekqqlkapeeevtfdf 240
 QY 942 YHKASRLGSSAMNIOMLLEKLGIPQLEGQKDMITWILQMDTVKRLKSKCLIVATTSG 1001
 DB 241 yhkasrlgssamniomlleklgipqlegqkdmittwilqmdftvkrlekscclivattsg 300
 QY 1002 SESIKNDEITHDKAEAKRRKAEAAARLHRQKIM 1034
 DB 301 sesikndeithdkeakerrkaeaaarlhrqkim 333

RESULT 7
 ID AAM25572
 ID AAM25572 standard; Protein: 250 AA.
 XX
 AC
 AC AAM25572;
 XX
 DT 16-OCT-2001 (first entry)
 XX
 DE Human protein sequence SEQ ID NO:1087.
 DE
 DE
 XX
 KW Human: cancer; HIV infection; human immunodeficiency virus;
 KW antinflammatory; antirheumatic; antiarthritic; immunosuppressive;
 KW antibacterial; endocrine; cardiant; central nervous system; virucide;
 KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
 KW antiaggregant; haemostatic; vulnery; antiulcer; osteopathic; eczema;
 KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;
 KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
 KW immunostimulant; gene therapy; antisenese therapy; vaccine; inflammation;
 KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
 KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
 KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
 KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
 KW allergic rhinitis; diabetes; multiple sclerosis; depression;
 KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
 KW neurological disorder.
 XX
 OS Homo sapiens.
 OS
 PN WO200153455-A2.
 PN
 XX 26-JUL-2001.
 PD
 XX
 PF 22-DEC-2000; 2000WO-US35017.
 PF
 XX 23-DEC-1999; 99US-0471275.
 XX
 PR 21-JAN-2000; 2000US-048725.
 PR
 PR 25-APR-2000; 2000US-0552317.
 XX
 XX (HYSE-) HYSEQ INC.
 PA
 PA Tang YT, Liu C, Drmanac RT;
 PI
 PI WPI; 2001-457603/49.
 XX
 DR N-PSDB; AAH99513.
 DR
 XX
 XX Isolated human polynucleotides encoding polypeptides, useful for the
 PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
 PT
 XX
 PS Claim 20; Page 220; 1217pp; English.
 PS
 XX AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
 CC AAM25963. The proteins can have activities based on the tissues and
 CC cells they are expressed in, such as: antinflammatory; antirheumatic;
 CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;
 CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;
 CC cardiovascular; antianaemic; antiaggregant; haemostatic; vulnery;
 CC antiulcer; osteopathic; dermatological; antiallergic; antiasthmatic;
 CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
 CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
 CC encoding them can be used in gene therapy, antisenese therapy and vaccine
 CC production. The proteins and polynucleotides are useful for screening for
 CC agonists or antagonists of a protein and for the treatment and diagnosis
 CC of disorders associated with the activity of a protein e.g. inflammation,
 CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
 CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
 CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
 CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
 CC osteoporosis, severe combined immunodeficiency, eczema, allergic
 CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
 CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
 CC neurological disorders.
 CC
 SQ Sequence 250 AA;

CC diagnosis, treatment and prevention of: (a) autoimmune diseases e.g.
 CC rheumatoid arthritis; (b) hyperproliferative disorders e.g. neoplasms
 CC of the breast or liver; (c) cardiovascular disorders e.g. cardiac
 CC arrest; (d) cerebrovascular disorders e.g. cerebral ischemia; (e)
 CC angiogenesis; (f) nervous system disorders e.g. Alzheimer's disease; (g)
 CC infections caused by bacteria, viruses and fungi; and (h) ocular
 CC disorders e.g. corneal infection. The polypeptides can also be used to
 CC aid wound healing and epithelial cell proliferation, to prevent skin
 CC aging due to sunburn, to maintain organs before transplantation, for
 CC supporting cell culture of primary tissues, to regenerate tissues and in
 CC chemotaxis.

XX Sequence 247 AA;

Query Match 3.2%; Score 295; DB 21; Length 247;
 Best Local Similarity 30.2%; Pred. No. 1.9e-16;
 Matches 74; Conservative 43; Mismatches 90; Indels 38; Gaps 9;

QY 1501 LKNGITPYLRCAALFHYLLGVTPPELHTNSAEGYSALCSYLS-LPTNLFLLFOEYWD 1559

Db 5 lqdfcplritsllqhlfg----edlpscgeeeefsvlasclgilpt-----fyg 52

QY 1560 TVRPLLORCAD--PAL---LNCCLKOKNTVVRYPR-----KRNLSLIELPDD 1600

Db 53 tehpffisasclwvvpafdiitqwcfeiksftrhaeqgkalliqeskwkplhllqipen 112

QY 1601 YSCLLNQASHFRCPRSADDERKHPVLCFCGALCSQNICCCQIVNGEEVGACIFHALHC 1660

Db 113 yntif-qyyhrktsvctkvpkdpavclvcgtfvelkigckq---qsycecvlshqnc 167

QY 1661 GAGCVFLKIRCRVVLVSGKARGCAYPAPYLDYGETDPLGRGNPLHLRSRYRKHL 1720

Db 168 gagtgifllinasviiiirgh-rfcilwsgvyldahgeedrldrrgkplyckerykvleg 226

QY 1721 VVQQH 1725

Db 227 qwish 231

RESULT 9

AA041592
 ID AAB41592 standard; Protein; 2096 AA.

XX AAB41592;

XX 08-FEB-2001 (first entry)

DE Human ORFX ORF1356 polypeptide sequence SEQ ID NO:2712.

XX Human; open reading frame; ORFX; detection: cytostatic; hepatotropic;
 KW vulnary; antipruritic; antiparkinsonian; nootropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antinflammatory;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antinflammatory disease; coagulation;
 KW thrombosis; contraceptive.

OS Homo sapiens.

XX W0200058473-A2.

PN 05-OCT-2000.

XX 31-MAR-2000; 2000WO-US08621.

PF

XX

XX

XX

XX

XX

PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.

XX (CURA-) CURAGEN CORP.

XX Shimkets RA, Leach M;

XX WPI; 2000-602362/57.

XX N-PSDB; AAC75801.

PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -

XX Claim 11; Page 1952-1956; 5507pp; English.

XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
 CC antipruritic; antiparkinsonian; nootropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiac; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 CC antithyroid; and antianemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 2096 AA;

Query Match 2.1%; Score 197.5; DB 21; Length 2096;
 Best Local Similarity 17.3%; Pred. No. 1.3e-06;
 Matches 259; Conservative 236; Mismatches 518; Indels 485; Gaps 60;

QY 117 IDPTCVLCMDCFQDSVHKHRYKMTSTGGGFCDCGDTFANKTGFCVYNHEPRAGTKE 176

Db 645 ldetvsvckka-qenmkqrhenethl-----eeqtsdlkmkkaelggqaavike 693

QY 177 ---NSRCPLENEVIVQARKIFPSVIVVEMTWEEKELEPPPELQIREKNERYCVLFND 233

Db 594 ahheatcrheee-----kkqlqvkleekthlqekrlqcheme-----731

QY 234 EHSYDHVIVSLQALDCELAELAAQHTTAIDKGRRAVAKAGY-----AACQEAKEIDK 287

Db 732 -----lkarlqag---asfgr-reglqssawteekvrglqleq---769

QY 288 SHSEN---VSQHPHVEVLHSEMAHOKFALRLGCSWNKIMSYSSDFRQFCQACLEEE 343

Db 770 fhqeqltslvkehkleekrlqleq-----feq 807

QY 344 PDSNPCLISRLMDAKLYKGARKILHELIFSSFFMEYKKLFAFEVYKYKLOREY 403

Db 808 metecnrttsqi---eaqfscdqkvtercsalqslqgryr-----qelkdleqq 856

QY 404 ISDDHRSISITALSVMQFTVPTLARHLIEBQNVISVITETLLEVLPE-YLDRNNKF--N 460

Db 857 reeksqvefekdelteqaeaqellketlkrkettstlvtqeremlektykdhlnsmvve 916

QY 461 FQGSYQDKLGRVAVICDLKYILISKPIWTER---LRMQFLEGFRSLKILTCMQGM-E 516

Db 461 FQGSYQDKLGRVAVICDLKYILISKPIWTER---LRMQFLEGFRSLKILTCMQGM-E 516

Db	917	rqqlldq-----ledlrnvsetqgllsdqlelkskhrlelreevl-cqggvse	967
Qy	517	EIRROGQHIEVDPOWEAAIAIOMLKNILLMFOWEACDEELL--LVAYKECHKAVMRC	574
Db	968	qlasrlrlrlemehqer-----qemskllamenlhkat--c	1003
Qy	575	STGPISSSTVVQSGHSLFTKSYKRVSDLVSIPLSRTLAGLHVLSRL-GAVSRLHE	633
Db	1004	eta-----dreaemsteisrlqskiemqq	1029
Qy	634	FVSPEDQVEVLVEPLRCLVLVAQVVAEMRRNGLSLISQVFFYQDVKCREEMYKDII	693
Db	1030	ats-----plsmilsggcqigve-----eevegdaIs	1056
Qy	694	MLQIGASMDPNKFLLLVLQR-YELAEAFNKTISTKDODLLIKOYNTLIEMLQVLIIVG	752
Db	1057	llkqgeqleengdvlslqraheqavkenvkmateisrlqgrlqkl--epglvmsscld	1114
Qy	753	ERYVPGVGNVTKEEVTMRREIHLILCIEPMPSAJAKNLPENNE---TGLEVINKVKAT	809
Db	1115	epateffgntaeqtepfllqnrktqvegtrhrvlsdleddevrdlsgtstssvqrqevk	1174
Qy	810	FKPFGVGHGVYELKDESLKDFNMVFFYHSYKTOHSKAEHMQKRRKOENKDEALPPPPP	869
Db	1175	ieeseasvegfelen-----seetrteswelkn-----	1204
Qy	870	EFCPAFSKVINLLNCDIMMYILRTVFERAIDTDSNLTGEMLOWAFHLLALGLLEKQO-	928
Db	1205	-----isllqqlmfmf-----adcdlasekqgelldfv--svlkkklk	1242
Qy	929	LOKAPEEYTFDYHKASRLGSSAMNIQMLEKLGIPQLEGQXDMITWILQDFVTVKRL	988
Db	1243	leripeasprykklyedvsrendclqealemmetrydealenkeltaevfrlqdelkkm	1302
Qy	989	-----REKSLII-----VATVSGSESTKND	1009
Db	1303	eevtetflsleksydevkieneelnvlrlqgkieklxtrawssgvtaaygkxslenle	1362
Qy	1010	ITHD-----KEKAERKRKA-----EAARLHRQIKMAQMSALOK-----NFI	1045
Db	1363	iepgnllqldtleecvprvrsvhvieeckqenqylegntllkxvkaheawlhgltl	1422
Qy	1046	ETHK-----LMYDNTSEMPKEDSIMEESTPAVDYSRIALG-----PKRGP-S	1089
Db	1423	qthqerprvqnqviieentt-lllgfdqhfghqataielektklqeltrklkervpil	1481
Qy	1090	VTEKEVLTILCOBQBEYVK-TENNAMVLISACVQKSTALTQHRKPIELSGEAL-----	1141
Db	1482	vkdkdvls--pgkkeeelkammhdlqipcsemqgkvellkyeseklqgenlsilrneittl	1539
Qy	1142	---DPLFMDPDPLAYGTYTSGCHVMHVAWCQYKFEAVOLS-----SQORIH	1184
Db	1540	needsi---snklqglngs-----qeenmwqtesvkgenaavlkmvnllkqliselklk	1591
Qy	1185	VDLFDLES-----GRYLC-----P	1198
Db	1592	nqqlldientelsqknsnpqeklqelnqlltemlqcqkekepgnsaleareceqfnlkeepe	1651
Qy	1199	LCKSICNVTIPI-----IPLQPOKINSEN---ADALAQLLTLARW-----TOTVLARI	1243
Db	1652	rckvqstlvslealsevklqthvqgenpllqdelmekmqhlhrpcpdlisnfgqklssv	1711
Qy	1244	SGYNIRHAKGENPIPIFFN---QGMGDSITFEHSILSGFGE--SSIKYSNSIK-----	1291
Db	1712	lsyneklkeaeiselncsvdklakslllehriaumkqeqkswehqhsasllqtlvasq	1771
Qy	1292	EMVILFATTIYRIGLKVPPDERDPVPMPLTWSTCAFTIQAIENTLLSGEGRPLFCALQNR-	1350
Db	1772	ekvqnledtqvnqlqmsrkmksdprv-----tqkeakealkqevmplhkkqlqnsv	1820
Qy	1351	-----OHNLKALMOFAVAQRITCPOVILIQKHLVRLSVLVPNTKSEDT	1394
Db	1821	xkswapeiathpsglhndqkrlrswdkldhlmneeqqllwqenerlqumv-gnltkael	1877

RESULT 10
AAM40027

ID AAM40027 standard; Protein; 2048 AA.

AC AAM40027:

DT 22-OCT-2001 (first entry)

XX Human polypeptide SEQ ID NO 3172.

KW	Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW	peripheral nervous system; neuropathy; central nervous system; CNS;
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemoaetactic;
KW	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW	leukemia.

OS Homo sapiens.

PN WO200153312-A1.

26-JUL-2001.

26-DEC-2000: 2000WO-US34263.

21-JAN-2000: 2000US-0488725.

PK ZJ-AFK-2000; 2000US-0552317;
PR 09-JUL-2000; 2000US-0598042;

PR 03-AUG-2000; 2000US-0653450.

PR 19-OCT-2000: 2000US-0693036.
PR 19-OCT-2000: 2000US-0693036.
PR 19-OCT-2000: 2000US-0693036.

FR 23 NOV 2007 1000Z
XX
XX
144: 0717103000Z

(HISE) HISEQ INC.

PI Yang H., Liu C., Asundi V., Wehrman T.
PI Wang J., Wang Z.

FI ZIRAO QA, ZIRAO F, GOODTICH
XX

DR WFI, ZOOI 442233/47.
DR N-PSDB: AAI59183.

PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
AA

Example 4: SEO ID NO 3172: 10078pp; English: XX

CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic.
CC

immunosuppressant and cytostatic activity, the polynucleotides are used in gene therapy. A composition containing a polypeptide or polynucleotide of the invention may be used to treat diseases of the peripheral nervous system, such as peripheral nervous injuries, peripheral neuropathy and localised neuropathies and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager Syndrome. Other uses include the utilisation of the activities such as: Immune system suppression, Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, cancer diagnosis and therapy, drug screening, assays for receptor activity, arthritis and inflammation, leukaemias and C.N.S disorders.

CC model: the sequence
CC specification.

Query Match 2.1%; Score 189.5; DB 22; Length 2048;
Best Local Similarity 18.1%; Pred. No. 5.9e-06;
Matches 268: Conservative 230; Mismatches 530; Indels 453; Gaps 65

SQ Sequence 1411 AA; Query Match 2.0%; Score 184; DB 17; Length 1411;
 Best Local Similarity 17.6%; Pred. No. 9.3e-06;
 Matches 193; Conservative 206; Mismatches 433; Indels 264; Gaps 40;

QY	176	ENSRCPNNEEVIVOARKIPFVSVIKYVVEMTWEEKELPELQIREKNERYCVLFNDHEH	235
DB	275	elagqpcevavvyqelgklkssvneltknqtitenllkkeqdytlkeekh-----nees	329
QY	236	HSYDHYIYSL-QRALDCELAEAQLHTTADKEGRRAVKAGAYAACQAKEDIKSHSENV	294
DB	330	vskknigatlhqldcdqqlgsrl-----sase-ts	359
QY	295	QHPLHVVLHSETMAHQKFALRGISNMNKIMSYSDPFQIFCOACLRREPDSNPCLISR	354
DB	360	lhrhive-lsekatqklkeelsevetkyghlkaefklqqg---reekehgqlgise	415
QY	355	LMLWDKLYKGARKILHELIFSFFMEMEYKKLFAMFYKYRQLOQKEYISDDH----	410
DB	416	inqhlsl-----leterqlgeahgrlkeqrqlsseklmdkeqqvadl	458
QY	411	SISTALSVM-----FTVPTLARHLI-----EQNVISVITETLL-----EVLPEY	452
DB	459	qlksrleeclkekvnstelhqldkqhdeqalqgsttakilreaqndleqvlrqi	518
QY	453	LDRNKF-NFGVYSQ-----DKLGRVY-----AVICDLK---YLISKPT	488
DB	519	gdqdkiqnleallqkskenisilekeredylakigagetavlnqlgeknhltlqe	578
QY	489	IWTERLRMQF-----LEQFRSFL-----KILTCMOGMEEIRVOGHIEVDP	530
DB	579	qlteklknqgeshkqaenlhqdvqeqkahllraaqdrvisletsvelnsqleskek--	636
QY	531	DWEAIIAQMOLNILMFOWCACDEELLVAYKECHKAVMRCSSTFSSTKTVVQSCG	590
DB	637	--vsqldiqikaktelllsaeaaaktagraqlqnlhdtagnalqdkqgelnkittqid	694
QY	591	HSLETKSRYRSE-----DLVSITHLPRLSRTLAGLHVRLSLGAVS---RLHFVSPED	641
DB	695	akiqdkhecsqleshkeyekyilslegkteeligkkleadslevkaskeqalgdlq	754
QY	642	V-EVLVEYPLRCLVLVAQVAE-----MWRNLGLSLISOVFYQDVVKCREMY	688
DB	755	qqqlntdlelatelskqliemekeivsstrldqkksalesikqkltkqe-----	806
QY	689	KDITIMIQIGASLMDPNPKFLLLVLORYELAEPNKTIS-----TKDQDLIKQYNLT	742
DB	807	eekqilqkdftelisgetk-----iqheelnnriqtvtelqkvkmekealmelstsv	861
QY	743	MLOVLIIVGER-----YYPGVGNVTKEEVTMRIEIHLCIEPMPHSAIAKNLPENN	796
DB	862	iskvsdsiksksfekengkgkaaiddlektckelkhglvg-----	904
QY	797	ETGLENVINKVATFKPGVSGHGVELKDLSLDNFNNFYVHYSKTOHSRAEHQMOKRR	856
DB	905	---mentlkqeeklk---slekeasqhqlkelmsgeqligaqntlknkeeeqqqlg	958
QY	857	ENKDEALPPPPFPFCAPFSKVINLNCDIMMYIL-RTPFERAIDTDSLNLWTGMLOMAF	915
DB	959	gninel-----kqsseqkkqiealqgelkavlqktenklqqq--ltqaqaelaa	1009
QY	916	HILAIGLLEEKOLOKAPAEVTFDYHKASRLGSAMNTQMLEKLKGIPIQLEGQKDMI	975
DB	1010	eekisvlqnykesqetfklqsdrygreseillatrqlkdvveekls-----laqed	1064
QY	976	T-----WILQMFDTVKRLREKS-----CLIVATTSGSESIKNDIEITHDK-	1014
DB	1065	snnrqignqnkligelkatleqsakeqqqlqerchkalqdigkeksikekelvneks	1124
QY	1015	-----EKAERKRKAEARLHRQKIMAQSALQC--NFIEETHKLMTONTSEMPGCDSI--	1065

Db 1125 klaeieeikcrqceitklneelkshkshlesikeitnlkdakqligdklelgqkadsika 1184
 QY 1066 -MEESTPAVS DYSRIALGPKRGPSVTEKEVLTCILCOEQEVEVK---IENNAMVLSACVQ 1121
 Db 1185 aveqe-----krnqgilkdgvk-----keeeelkefeieakhseike 1224
 QY 1122 KSTALTQHRCGPIELS 1137
 Db 1225 kevgmkkheeneaklt 1240
 RESULT 12
 AAY77575
 ID AAY77575 standard; Protein; 2442 AA.
 XX AC AAY77575;
 XX DT 08-MAY-2000 (first entry)
 XX DE Human cytoskeletal protein (HCYT) (clone 1655208).
 XX KW Human cytoskeletal protein; HCYT; cell proliferation; immunological;
 KW reproductive; nervous disorder; cytostatic; immunosuppressive; anti-HIV
 KW anti-diabetic; antiarteriosclerotic; dermatological; anti-inflammatory;
 KW anti-infectivity; vasotropic; cardiac.
 OS Homo sapiens.
 XX WO2000006730-A2.
 PN 10-FEB-2000.
 PD 30-JUL-1999; 99WO-US17167.
 PF 31-JUL-1998; 98US-0155185.
 PR 04-AUG-1998; 98US-0160081.
 PR 19-AUG-1998; 98US-0155228.
 XX (INCY-) INCYTE PHARM INC.
 XX Bandman O, Tang YT, Yue H, Corley NC, Guegler KJ, Azimzai Y;
 PI Patterson C, Lal P, Baughn MR;
 XX WPI: 2000-195297/17.
 DR N-PSDB; AA258980.
 XX Human cytoskeletal proteins useful for diagnosing, treating preventing
 PT cell proliferative, immunological, reproductive, developmental and
 PS nervous disorders
 XX Claim 1; Page 71-76; 84pp; English.
 CC The invention provides human cytoskeletal proteins (HCYT) and nucleic
 CC acids encoding the proteins. The HCYT polypeptides can be expressed by
 CC standard recombinant methodology. The HCYT polypeptides, modulators and
 CC antibodies are useful for treating or preventing a disorder associated
 CC with decreased and increased expression or activity of HCYT in mammals.
 CC The polypeptides are also useful for diagnosing HCYT activity disorders
 CC such as cell proliferative, immunological, reproductive, developmental
 CC and nervous disorders. Sequences AAY77569-576 represent HCYT
 CC polypeptides.
 XX Sequence 2442 AA;
 SQ
 Query Match 2.0%; Score 182; DB 21; Length 2442;
 Best Local Similarity 20.9%; Pred. No. 3.4e-05;
 Matches 256; Conservative 187; Mismatches 478; Indels 304; Gaps
 QY 207 WE--EKELPPELOIREKNERYCVCVLENDEHHSDHYVLSQRLDCELAE-----AQ 257
 Db 88 wenveepnl-dellvrleeegqrcceslae-----vntqirlhmekadvnvkairad 137

QY	174	IKENSRCLNEEVIVOAARKTTPSPVIKYVEMTWIWEKEKLPPEL-----QIREKNERYYC	228
Dd	757	veene--alreevi-----llseklspseverlrkeiqksee--	794
QY	229	VLFNDEHHSDHYISLORALDCELAEAQLHTTAIDKEGRRAVKAGAYACQAKEDI--	286
Dd	795	-----hiits-----ekdklfsevnhkesr---vqglleeigtkddlat	831
QY	287	-KSHSENVSOHPHVEVLHSEIMAHQKFALRLGSGWN-KIMSYSDDFRQIFCQACLREEP	344
Dd	832	tqsnystdqefnftlmdfegkykmvleenernmqeivnliskeaqkf-----	881
QY	345	DSENPCILSRLMDAKLYKGARKI--LHE-----LIFSFF	378
Dd	882	dsslgalktelsyktqlgektevqerlnemeqlkeqlenrdspqtverehtlitekl	941
QY	379	FMEWEYKKLFAMEFVYYKQLOK-----EYTSDDHDR-SISITALSVOMFTVPTLARH	430
Dd	942	qqtleevktltg-kadlkqqlgeslgierdqlksdhdtvmnmidtgelrlnaleslkqh	1000
QY	431	LIEEQNVISVITETTLLEVLPE--YLDRNNKFNFGYSODKLGRVYAVICDLKVILISK--	486
Dd	1001	----getintlskiseevsrnlhmeentgetkdfeqqkmvg-----dkqdleaaknt	1050
QY	487	PTWTERLRMQLEGFSFLKILTQMGMBEIRRVGOHIEVDPDWEAATAIQMLK---	543
Dd	1051	qtltaadvkdnelleqr---kifsligeknelqgm-----esviaekeqiktdl	1097
QY	544	--NILMFQEWCADEBELLVA--YECHKAVMRCSTSFSSSKTVVQSCG-----HSL	593
Dd	1098	kenientien---qeelrlgdellkkqevaageknaikkegelrtsrtcdrlaeveekl	1153
QY	594	ETKSRYRSE--DLVSIIHLPLS-----RTLAGHLVLSRLGAVSRLH	632
Dd	1154	kelsqqlkeqqqlnvqeemseqbkineienknelnkeltlehmeterlelaqkin	1213
QY	633	EYVSFEFOVELVLEYRLCLVLVAQVAMWRNRGLSLISQYFYQDVKCREMYDKOI	692
Dd	1214	e--nyee-----vksitke--rkvlkelqksfeterdhrlrgyreiea	1252
QY	693	IMLIQIGASLMDPNKFFLLVLORY-ELAEAFNKTISTR-----DQDLIKOYNLTLEMLQ	745
Dd	1253	tgigtkeel-----kiahlnlkehetidelrrsvsektaqiintdgleshtklgee--	1305
QY	746	VLIVYIGERYVPGVGNVTKEEVMRIIHL-----LCIEPMPHSAIAKNLPENE	794
Dd	1306	ipvlheeqlpnvkvsqetqmnelellteqstktdsttlariemerlrinekfqeq	1365
QY	795	NNETGLEVINKATPKPGVGHGYVELKDESCKDFNMIFYH-----YSKTQHSKAEHMQ	850
Dd	1366	eeikstkerdnltkike-----alevhdklke-----hiretlakigesqsqkeq	1412
QY	851	KRRKQENKDEALPPPPPFCFAPFSKVINLNCDIMVIL-----RTV-----	894
Dd	1413	slnmkekdnctkiwsemegfkpkdsaa---llrielemjglskrlqeshdemksvakekd	1469
QY	895	----FERADTDNSLMTGEMLO-MAPHILALGLLEEKOOLOKA-----PEEEVTFDYHK	944
Dd	1470	dlarlqevigesdqikenikelvakh-----letecelkvahccлкеqeteinelrvn	1523
QY	945	ASRLGSANNIOMLLE-----CLKGIPQLBSQKDM---ITWILQMFDTKRLREKSCL	994
Dd	1524	lseketistiqkleaindklnkqlkeiyekееeqinikisevqekvnelkafkehr--	1581
QY	995	IVATTSGSESINKD--EITHD-----KEAERKRKAEARLHR-----QKI	1033

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 1, 2002, 12:32:35 ; Search time 2649.68 seconds
(without alignments)
25582.097 Million cell updates/sec

Title: US-09-724-126A-1
Perfect score: 6308
Sequence: 1 gccagaattgcgcagcaggg.....aatttgattgggtgtttt 6308

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_eston:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_estl:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	674.6	10.7	682	10	AI929033
2	656.6	10.4	756	11	BG334574
3	637.4	10.1	641	10	AI361043
4	627.2	9.9	797	11	BG862813
5	573.2	9.1	782	11	BI086469
6	557	8.8	565	11	BF063405
7	512.2	8.1	606	10	BE589438
8	479.8	7.6	624	11	BG219270
9	477.8	7.6	481	10	AI192195
10	473.4	7.5	583	10	AW971391
11	471.6	7.5	478	10	AA401319
12	454	7.2	522	10	AW291190

13	449	7.1	522	11	BG382648
14	448	7.1	520	11	BG382624
15	410.8	6.5	486	10	BE550873
16	354.8	5.6	925	10	BE573578
17	352	5.6	374	10	AA400279
18	351.4	5.6	435	10	AA489271
19	330	5.2	394	11	BF774117
20	318	5.0	481	10	AI187306
21	307.2	4.9	397	10	AW311960
22	301.8	4.8	504	10	AI980640
23	266.2	4.2	327	10	BE654236
24	265.2	4.2	523	11	EG979647
25	263.4	4.2	300	11	217892
26	252.6	4.0	1079	11	BG292980
27	249	3.9	562	10	AI504731
28	243.6	3.9	372	11	BF542537
29	242.8	3.8	505	11	EG364916
30	241	3.8	325	10	AI615529
31	237	3.8	291	10	AV225341
32	235.8	3.7	396	11	BE930958
33	232.8	3.7	256	10	AI921294
34	231.4	3.7	278	10	BE077143
35	230.6	3.7	323	11	H33916
36	223.6	3.5	745	10	AW976158
37	222	3.5	555	11	BE873236
38	217	3.4	685	10	AI693180
39	209	3.3	313	10	AV168252
40	208.6	3.3	632	11	BG625558
41	206.8	3.3	221	10	AA507138
42	204	3.2	1148	11	BF164318
43	201	3.2	620	10	AW702134
44	200.4	3.2	435	11	BE930879
45	200	3.2	711	10	AI646734

ALIGNMENTS

RESULT 1

AI929033 682 bp mRNA EST 23-AUG-1999
LOCUS au64c10.v1 Schneider fetal brain 00004 Homo sapiens cDNA clone
DEFINITION IMAGE:2519538 5' similar to TR:070481 070481 UBIQUITIN-PROTEIN
LIGASE E3 COMPONENT N-RECOGNIN ; mRNA sequence.

ACCESSION AI929033 GI:5664997
VERSION AI929033.1
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 682)
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,
J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
White,Y., Wylie,T., Waterston,R. and Wilson,R.

TITLE WashU-NCI human EST Project
JOURNAL Unpublished (1997)
COMMENT Other_ESTs: au64c10.x1
Contact: Wilson RK

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Possible reversed clone: similarity on wrong strand
Seq primer: -40RP from Gibco
High quality sequence stop: 450.
Location/Qualifiers
1. 682
/organism="Homo sapiens"


```
/db_xref="taxon:9606"
/clone="IMAGE:2519538"
/clone_lib="Schneider fetal brain 00004"
/sex="male"
/tissue_type="frontal lobe"
/dev_stage="5 months post-conception"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescript SK (Stratagene);
Site 1: SstI; Site 2: XhoI; Double-stranded cDNA was
prepared from human fetal brain tissue. 5' and 3'
adaptors were used in cloning as follows: 5' adaptor
sequence:
5'-GAGAGAGAGAGAGCTCAAGGATCCTTAATTAATTAATCCGCCCCGCC-3'
and 3' adaptor sequence:
5'-GAGAGAGAGAGCTCAGTGTCTTTT-3'. The library was
size-selected for >0.5 kb inserts and has an average
insert size estimated at 1.2 kb. This library was
constructed using the CAP-trapper method for full-length
enrichment and has not undergone amplification. Library
was constructed by Dr. Claudio Schneider (LNCIB-Area
Science Park, Trieste, Italy)."
BASE COUNT 178 a 155 c 166 g 182 t 1 others
ORIGIN

Query Match 10.7%; Score 674.6; DB 10; Length 682;
Best Local Similarity 99.3%; Pred. No. 3.9e-147;
Matches 677; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 5261 aactccgctgaggaactgcataccaattctgcagaagagagacagtcactctgtag 5320
Db 1 AACTCCGCGCTGAGGAAGTGCATACCAATCTGCAGAAGGAGAGTACAGTGCACCTGTAG 60

Qy 5321 ctatctatttaactacaattgttctctctctcaggaataattgggatactgttaag 5380
Db 61 CTATCTATCTTTTACCTACAAATTTGTCCTCTCTCTCAGGAATATGGGATCTGTAG 120

Qy 5381 gccctgtccagagtggtgtgcagatccctgccttacttaactttgagcaaaaaa 5440
Db 121 GCCCTGTCTCAGAGGTGGTGTGCAGATCCGCTTACTTAACCTGTTTGAAGCAAAAAA 180

Qy 5441 caccgtgtcaggtaccctgaagaaagaaatagttgtatagacttctctgatgactatag 5500
Db 181 CACCGTGTCTCAGGTACCTCAGAAAGAAATAGTTGTATAGAGCTTCTGTATGACTATAG 240

Qy 5501 ctgctctgatacaagcttctcattcaggtgcccaggtctgcagatgatgagcgaaa 5560
Db 241 CTGCTCTCTGAATCAAGCTTCTCATTTTCAGGTGCCCGCTCTGCAGATGATGAGCGAAA 300

Qy 5561 gcatcctgtcctctctctctctgtgggctatactatgttctcagaacatttgcctga 5620
Db 301 GCATCCTGTCTCTGCTTTTCTGTGGGGCTATACATATGTTCTCAGAACATTTGCTGCCA 360

Qy 5621 ggaattgtgaacggggaaggttgagcttgcatttctcagcacttcaactgtggagc 5680
Db 361 GGAATTTGTAACGGGGAAGGTTGGAGCTTGCATTTTTCACGCACCTCACTGTGGAGC 420

Qy 5681 cggagctgtcatttctctaaatacagagaatgccgagtggtcctggttgaaggttaagc 5740
Db 421 CGGAGCTGTGCATTTTCTCTAAATAATCAGAGAAATGCCAGTGGTCTGTGTAAGGTAAAGC 480

Qy 5741 cagaggctgtgcctatccagctccttaacttgatgaataatggagaaacagaccctggcct 5800
Db 481 CAGAGGCTGTGCCCTATCCAGCTCCTTACTTGGATGAATATGGAGAACAGACCCCTGGGCT 540

Qy 5801 gaagaggggcaaccccttatttctcgtgagcggtatcggaagctccatttgctgtg 5860
Db 541 GAAGAGGGGCAACCCCTTCTATCTCTGTGAGCGGTATCGGAAGCTCCATTTGGTCTG 600

Qy 5861 gcaacaacactgcattatagaagagattgttaggagcgaagactcaatcagatgttatt 5920
Db 601 GCAACAACACTGCATTATAGAAGAGATTGCTTAGGAGCCACAGAGACTAATCAGATCTATT 660
```

```
Qy 5921 tggattcaactggcagttactg 5942
Db 661 GGGATTCAACCTGCAGNACTG 682

RESULT 2
LOCUS BG534574 756 bp mRNA EST 03-APR-2001
DEFINITION 602553425F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4663182 5',
mRNA sequence.
ACCESSION BG534574
VERSION BG534574.1 GI:13526116
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 756)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLC1465 row: a column: 07
High quality sequence stop: 751.
Location/Qualifiers
1..756
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4663182"
/lab_host="NIH_MGC_77"
/note="Organ: lung; Vector: pDNR-LIB (Clontech); Site 1:
SfiI (ggccgctggccc); Site 2: SfiI (ggccattggccc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CAGGCCATATGAGGC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGGCGGCGGACATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.9
kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH_MGC Library."
BASE COUNT 274 a 138 c 152 g 192 t
ORIGIN

Query Match 10.4%; Score 656.6; DB 11; Length 756;
Best Local Similarity 95.9%; Pred. No. 6.5e-143;
Matches 730; Conservative 0; Mismatches 19; Indels 12; Gaps 5;

Qy 2836 ggtatgaacttgcggaggcttttaacaagaccatattacaaagaccagattgatta 2895
Db 1 GGTATGAACCTGCCGAGGCTTTTAAACAAGACCATTATCTACAAAAGACCAGGATTGATTA 60

Qy 2895 aacaataataacactaataagaagaatgctcaggctccatctatattgtgggtgagc 2955
Db 61 AACAATATATACACTAATAGAAGAATGCTTCAGGTCCTCATCTATATTGTGGGTGAGC 120

Qy 2956 gtatgtacctggagtggaatgtgaccaaagaaggtcacaatgagagaaatcattc 3015
Db 121 GTATGTACTGGAGTGGGAATGTGACCAAGAAGAGGTGCACANTGAGAGAAATCATTC 180

Qy 3016 acttgccttgcattgaaccatgccacacagtcgcatcgtcccaaaatttacctgagaatg 3075
Db 181 ACTTGCTTGTTCATTGAACCCATGCCACACAGTGCCTATGCCAATAAATTTACCTCAGAATG 240
```

QY	3076	aaataatgaactgcttagagaatgctataaacaagaatgagccacatttaagaataccag	3135
Db	241	AAAATAATGAATGAACTGGCTTAGAATGCTATAAACAAAGTGGCCACACTTTAAAGAAACGAC	300
QY	3136	gtgtatcagccatgagttatgaactaaagaatgaatcactgaagaagcttcaatgt	3195
Db	301	GTGTATCAGCCATGAGTTTATGAACATAAAGATGAATCACTGAAAGACTTCATATGT	360
QY	3196	acttttattcattactccaaacccagcatagcagggtgaacatatgcagagaagaaaagga	3255
Db	361	ACTTTTATCATTACTCCAAACCCAGCATAGCAAGGCTGAACATATGCAGAGAAAGGA	420
QY	3256	gaaacaagaatacaagaatgaag-----cattgcgcgcacaccacctctctgaattct	3309
Db	421	GAAAACAAAGAAAAAAGATGAAGGTAAAAACATTCGCGACCAACCACCTCTCTGAATTCT	480
QY	3310	gccctgttttcagaaagtgaataacccttcaactgtgatat-catgatgtacattctc	3368
Db	481	GCCCTGCTTTCAGCAAGTGATTAACTTCTCAACTGTGATATCCATGATGATCATTTCTC	540
QY	3369	aggacgtatttgagcgggcaatagacacagattcttaactgtgaccgaaggatgctc	3428
Db	541	AGGACCGTATTTCAGCGGGCAATAGACACAGATTCTAATTGTGTCACCGAAGGATGCTC	600
QY	3429	caaatgcttttcatacttctggcattggtgttactagaagaagcaacagcttcaaaaa	3488
Db	601	CAAAATGTTTTTCATATTCTTCGGCATTTGGTCTTACTAGAGAAGCAACAG-TTCAAAAA	659
QY	3489	gctctgaagaagaagtaaacatttgacttttataatgaagcttcaagattgggaagtcca	3548
Db	660	GCTCCTGAAGAAGAAGTAACATTGACTTTTATCATAGGCTTCAAGATTGGGAAGTTCA	719
QY	3549	gccatgaatacaaatgcttttggaataacatacaaaaggaat	3589
Db	720	G-CATGAATATACAA---TCGTTTGGAAAAATCAAGGATT	756
RESULT 3			
AI361043/c			
LOCUS	AI361043	641 bp	mRNA EST 15-FEB-1999
DEFINITION	qy03f11.x1 NCI CGAP_Brn20 Homo sapiens cDNA clone IMAGE:2010957 3' similar to TR:P91133 P91133 SIMILAR TOS. CEREBISIAE		
ACCESSION	UBIQUITIN-PROTEIN LIASE E3 COMPONENT SP:P19812.; mRNA sequence.		
VERSION	AI361043		
KEYWORDS	AI361043.1 GI:4112664		
SOURCE	EST.		
ORGANISM	human.		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 641)		
JOURNAL	NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .		
COMMENT	National Cancer Institute / National Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project (CGAP/BTGP), Tumor Gene Index		
	Unpublished (1998)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgapbs.r@mail.nih.gov		
	Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D., Ph.D.		
	cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.		
	cDNA Sequencing by: Greg Lennon, Ph.D.		
	DNA Sequencing by: Washington University Genome Sequencing Center		
	found through the I.M.A.G.E. Consortium information can be		
	www.bio.llnl.gov/bbrp/image/image.html		
	Insert Length: 1083 Std Error: 0.00		
	Seq primer: -400P from Gibco		
	High quality sequence stop: 453.		
FEATURES	Location/Qualifiers		
source	1..641		
	/organism="Homo sapiens"		

```

/db_xref="taxon:9606"
/clone_image=2010957"
/clone_lib="NCI_CGAP_Brn23"
/tissue_type="glioblastoma (pooled)"
/lab_host="DH10B"
/notes=Organ: brain; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
TGTTCACCACTCTCAAGTCGAGCGCGCATATCTTTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
Library is normalized, and was constructed by Bento
Soares and M.Fatima Bonaldo."
BASE COUNT      174 a   155 c   135 g   175 t
ORIGIN
Query Match      10.1%; Score 637.4; DB 10; Length 641;
Best Local Similarity 99.5%; Pred. No. 2e-138;
Matches 638; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 5457 cctgagaagaatagtttgtagagagcttctctgatactatactagctcctctgaatcaa 5516
|||||
Db 641 CCTAGAAAAAATAGCTTTGTATAGAGCTTCCTGATGACTATAGCTGCCTCTGAATCCA 582
|||||
QY 5517 gcttctcattcaggtgcccaagctctgcagatgatagcgaaagcatctctgctctgc 5576
|||||
Db 581 GCTTCTCATNTCAGTGCCTCACGGCTCTGCAGATGATGAGCGAAGCATCTCTGCTCTGCG 522
|||||
QY 5577 ctttctgtggggctatactatgtttctcagaacaattgtctgcaggaaattgtgaaacggg 5636
|||||
Db 521 CTTTCTGTGGGCTATACATATGTTCTCAGAACATTTGCTGCCAGGAAATGTTGAACGGG 462
|||||
QY 5637 gaagaggttgagcttgcatttttcaagcacttcaactgtggagccgagctgcatcttc 5696
|||||
Db 461 GAAGAGGTTGAGCTTGCAATTTTTCAGCGCACTTCACCTGTGGAGCGGAGTCTGCATTTTC 402
|||||
QY 5697 ctaaaaatcagagaatgccagtggtcctggttgaagtaaaagccagagcctgtgcctat 5756
|||||
Db 401 CTAAAAATCAGAGAAATGCCGAGTGGTCTCTGTTGAAGGTAAAGCCAGAGGCTGTGCCAT 342
|||||
QY 5757 ccagctccttacttggatgaatatggagaacacagaccctggcctgaagagggggcaacccc 5816
|||||
Db 341 CCAGCTCCTTACTTTGGATGAATATGGAGAAACAGACCCTGSCCTGAAGAGSGGCAACCCC 282
|||||
QY 5817 cttcatttatctctgagcgggtatcgagaagctccatttgctgtggcaacaacactgatt 5876
|||||
Db 281 CTTCAATTTATCTGCTGAGCGGTATCGGAAGCTCCATTTTGTCTGGCAACACACTGCAT 222
|||||
QY 5877 atagaagagattgctagagaccgaagacactaatcagatgttatttgattcaactggcag 5936
|||||
Db 221 ATAGAAGAGATTGCTAGAGCCAGAGACATAATCAGATGTTATTGTGATTCAACTGGCAG 162
|||||
QY 5937 ttactgtgagcttcaactctgcctcaagacaatcacaaaatgacacagtagtaagggctg 5996
|||||
Db 161 TTACTGTGAGCTCAACTCTGCCTCAAGACAATCACAAATGACGACAGTAGTAAGGCTG 102
|||||
QY 5997 attcaaatattgaaaaactttctgagggctgggaaagtattggagggctctttgtccca 6056
|||||
Db 101 ATTCAAAATTATGAAAACCTTCTGAGGGCTGGGAAAGTATTGGAGGGCTCTTTTGTCTCCA 42
|||||
QY 6057 tgtccaggttcacttaacataataataattttttaatgga 6097
|||||
Db 41 TGTCCAGGTTCACTTACATCAATAAATAATTTCTTAATGGA 1
|||||

RESULT 4
BG862813 797 bp mRNA EST 29-MAY-2001
LOCUS BG862813 602790074F1 NIH_CGAP_Mam4 Mus musculus cDNA clone IMAGE:4934370 5'
DEFINITION mRNA sequence.
ACCESSION BG862813

```